Middle Atmosphere Program

HANDBOOK
FOR MAP
VOLUME 13

Edited by
R.A. Vincent

ICSU

International Council of Scientific Unions

SCOSTEP

Scientific Committee on Solar-Terrestrial Physics
K. D. Cole, President
J. G. Roederer, Vice President
C. H. Liu, Scientific Secretary

MAP ORGANIZATION

MIDDLE ATMOSPHERE PROGRAM STEERING COMMITTEE
S. A. Bowhill, SCOSTEP, Chairman
K. Labitzke, COSPAR, Vice Chairman
C. H. Liu, SCOSTEP, Secretary

R. D. Bojkov, WMO	L. R. Megill, IUGG/IAGA
A. D. Danilov, COSPAR	T. Nagata, SCAR
J. C. Gille, COSPAR	R. G. Roper, IAMAP
A. H. Manson, SCOSTEP	P. C. Simon, IAU
I. Hirota, IUGG/IAMAP	J. Taubenheim, IUGG/IAGA
J. W. King, URSI	T. E. VanZandt, URSI

M. Wada, IUPAP

MAP STANDING COMMITTEES
Data-Management -- G. Hartmann and I. Hirota, Co-Chairmen
Dynamics Calendar -- T. E. VanZandt, Chairman
Publications -- C. F. Sechrist, Jr., Chairman

MAP STUDY GROUPS
MSG-5   Ions and Aerosols, F. Arnold and M. P. McCormick, Co-Chairmen
MSG-6   Scientific Aspects of an International Equatorial Observatory,
        S. Kato, Chairman
MSG-7   Penetration of Solar Radiation into the Atmosphere, J. E.
        Frederick, Chairman
MSG-8   Atmospheric Chemistry, G. Witt, Chairman
MSG-9   Measurement of Middle Atmosphere Parameters by Long Duration
        Balloon Flights, J. E. Blamont, Chairman

APPROVED MAP PROJECTS

	Coordinator			Coordinator
AMA:	T. Hirasawa		GOSSA:	M. P. McCormick
ATMAP:	J. M. Forbes		GRATMAP:	M. A. Geller
CAMP:	G. Witt		MAE:	N. C. Maynard
CLIMAT:	J. M. Russell		MSTRAC:	P. K. Rastogi
DYNAMICS:	K. Labitzke		OZMAP:	D. F. Heath
GLOBMET:	R. G. Roper		SSIM:	P. C. Simon
GLOBUS:	D. Offermann		WINE:	U. von Zahn

MAP REGIONAL CONSULTATIVE GROUP

Europe   M. L. Chanin, Chairman

M I D D L E

A T M O S P H E R E

P R O G R A M

HANDBOOK FOR MAP

Volume 13

Ground-Based Techniques

Edited by

R. A. Vincent

November 1984

N85-17452 #
thru
N85-17468 #

Page intentionally left blank

FOREWORD

A wide range of both active and passive techniques has been developed for remote sensing of the middle atmosphere from the ground. A significant fraction of the electromagnetic spectrum is utilized, ranging from the ultraviolet to low-frequency radiowaves, which means that many important parameters of the neutral and ionized atmosphere can be studied. The ability of many ground-based techniques to provide almost continuous monitoring of the state of the atmosphere, often with excellent time and height resolution, means that they are playing an important role in the Middle Atmosphere Program.

Improvements to existing methods and the development of new techniques has been especially rapid in the last decade and it is often difficult to remain abreast of the latest innovations in not only hardware but also methods of analysis. The advent of MAP has itself stimulated much new work and it is therefore appropriate that a Handbook for MAP be specifically devoted to providing technical information. To this end recognized experts were invited to provide chapters on their specialities and covering such aspects as instrumentation, data analysis and present and future capabilities. Due to reasons of space it has not been possible to cover all topics exhaustively, but it is hoped that the information provided will encourage others to take up one or more of these techniques.

Finally, I would like to thank the contributors for the time and effort they gave. Their cooperation was freely given and fully in the spirit of MAP.

R. A. Vincent

Page intentionally left blank

GROUND-BASED TECHNIQUES

CONTENTS

1. GROUND-BASED MEASUREMENTS OF MESOSPHERE TEMPERATURES BY OPTICAL MEANS

J. W. Meriwether, Jr.

Space Physics Research Laboratory
University of Michigan
Ann Arbor, MI 48109

ABSTRACT

The determination of the rotational temperature of the molecular emissions of OH or O_2 offers a means of observing from the ground the mesosphere temperature at about 90 km throughout the night. We describe the techniques and optical instrumentation developed to carry out these measurements. In comparison with other methods for the observation of this important geophysical quantity the optical technique is inexpensive and readily applicable for automatic operations over a long time period.

INTRODUCTION

The temperature of the mesosphere is an important physical parameter of vital interest to studies on D region ion chemistry and mesospheric dynamics. The chemical reaction rates for the water cluster ions involving NO^+ within this region are strongly temperature-dependent (ARNOLD et al., 1980; CHAKRABARTY and CHAKRABARTY, 1979; FORBES, 1982). It appears that the unusual behavior observed in ion composition measurements can now be explained by modifications of the reaction path involving water cluster ions that occur with changes in the mesosphere temperature (CHAKRABARTY and CHAKRABARTY, 1978). Hence, the details of the phenomenology of various mesospheric processes can be better understood by looking first at the temporal behavior of the temperature and considering how the observed changes could be related to the phenomenon in question.

The largest annual variation in mesospheric temperatures is found within the high latitude regions where the highest values are found in the winter and the lowest values, perhaps the lowest to be found within the Earth's atmosphere, during the summer. The annual variation may be as large as one hundred degrees. Figure 1 shows the distribution of mesospheric temperatures as a function of latitude and season (from FORBES, 1982, based on AFGL compilations by Cole and Kantor). The reasons for this variation contrary to physical intuition relate to the transport of atomic oxygen from the summer hemisphere to the winter hemisphere (COGGER and MURPHREE, 1980; COGGER et al., 1981; ISMAIL and COGGER, 1982). These two extremes in the range of the mesosphere temperature at high latitudes appear to be the key element in understanding the phenomenon of the winter anomaly (BEYNON et al., 1965; OFFERMANN et al., 1981; SOLOMON et al., 1982) and noctilucent clouds (WEBB, 1965; HAURWITZ and FOGLE, 1969; DONOHUE et al., 1972; ARNOLD, 1980; BJORN and ARNOLD, 1981; TURCO et al., 1981; GADSDEN, 1982).

There are also dynamical effects in the mesosphere introduced by gravity waves and tidal processes. The thermal signature of these short-lived dynamical events may be detected by looking for mesospheric thermal fluctuations of a few degrees with periodicities less than several hours in the case of the gravity waves (VISCONTI et al., 1971; KIEFFABER et al., 1972; MERIWETHER, 1975; NOXON, 1978; MOREELS et al., 1977; WEINSTOCK, 1978; TEPLEY et al., 1981) and with periodicities of a few hours in the case of tidal variations (PETITDIDIER and TEITELBAUM, 1977; MERIWETHER, 1979).

The sensitivity of the mesosphere ion chemistry to the kinetic temperature of this region, therefore, creates a requirement for measurements of mesospheric

ZONAL MEAN TEMPERATURES
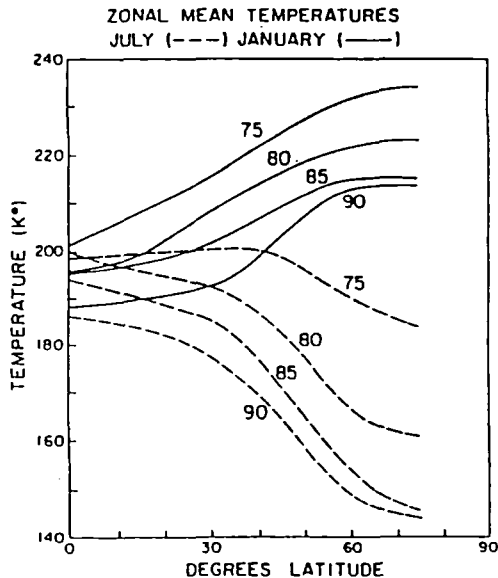JULY (- - -) JANUARY (———)

Figure 1. Latitudinal distribution of zonal mean temperatures for January and July for several mesosphere heights (from FORBES, 1962).

temperatures as an important part of any experimental study concerning the mesosphere. There are two ways to satisfy this need: either through "in situ" or through remote sensing measurements. The experimental task of the first in the mesosphere is quite difficult. This region is the regime of transition between the stratosphere where complete mixing prevails and the lower thermosphere where only binary collisions occur and diffusive equilibrium holds for individual species. Therefore, the normal means of measuring the temperature by "in situ" methods involving pitot tubes, probes, or falling spheres are greatly complicated by the various corrections needed in the data analysis.

An alternative option that we review in this paper follows the lead of techniques derived from the newly developed field of remote sensing. Visible rotational lines selected from molecular bands emitted by the terrestrial nightglow are isolated by means of low resolution optical instrumentation and their intensities measured in the relative sense. The magnitude of the mesosphere temperature in the direction observed may then be computed from the observed ratio of intensities.

Examples of the optical instrumentation used in the literature are the spectrograph utilizing photographic or video recording, the grating spectrophotometer, the field-widened Michelson interferometer, and the Fabry-Perot interferometer (either solid state such as in interference filters or as an etalon plate assembly with interchangeable spacers).

These measurements are made from the ground for any direction and analyzed to yield a temperature that is a quantity averaged over the volume emission profile along the instrumental line of sight. It is assumed in this method that the kinetic temperature of the mesopause may be determined from the measurement of the rotational temperature of mesospheric molecular emissions. For OH the radiative lifetime falls between the extremes of 3 milliseconds for the $v = 9$ vibrational state to 63 milliseconds for the $v = 1$ state (LLEWELLYN and LONG, 1978; WARE, 1980). In contrast the collision frequency within the mesosphere gives an intercollision time of about 0.1 millisecond, sufficiently short for the thermalization of the molecule after several collisions (KILLEEN and HAYS, 1981). This assumption has never been checked directly. The experiment would

be straightforward as what is needed is a comparison of the kinetic temperature as determined by a Fabry-Perot interferometer with the rotational temperature of the same band as determined by a lower resolution instrument.

Furthermore, there is the report in the literature by GATTINGER and VALLANCE-JONES (1972) of an observation of a non-equilibrium rotational population distribution for OH emissions. Their work suggests that there are times when the non-thermal distribution may persist over several hours for high rotational levels of a particular rotational band. Gattinger and Vallance-Jones argued that for sufficient collision frequencies thermal equilibrium would prevail, but for cases when the OH layer lies higher than the normal height of 85 km, then laboratory measurements indicate formation of the excited state of OH with a high degree of rotational excitation would be likely. Infrared measurements would be necessary for a study on this question, because only in the infrared where the OH emissions are very intense is it possible to apply instrumentation sufficiently sensitive to the weak emissions involving high K values of the P branch of the OH band. This matter looks to be serious only for the states of high K values, which require more time for equilibration with the medium.

For molecular oxygen, the lifetime is about 12 seconds (NOXON, 1970). Hence, collisions are sufficiently numerous to ensure thermal equilibrium. The normal height for the O_2 airglow layer is about 97 km (NOXON, 1978; WITT et al., 1979).

OPTICAL EMISSIONS OF THE MESOSPHERE

A feature of the transition character of the mesosphere is the formation of airglow layers generated by the production of metastable states through chemical reactions involving the photochemistry of this region (BATES, 1981, 1982) or related to the species added to the chemical population of the mesosphere by deposition of meteors (GADSDEN, 1969, 1970). The height of these layers lies within the mesopause region of 80 to 100 km, where the temperature profile displays a local minimum in temperature.

The strongest atomic airglow emission is emitted at the wavelength of 5577 Å, radiated by the atomic oxygen species, $O(^1S)$, at a height of 97 km (DONOHUE et al., 1974). Other mesospheric emissions of significant brightness and molecular in origin result from reactions involving metastable species of atomic oxygen (BATES, 1981). Chief among these are the bands of the atmospheric system with band origins located near 7600 Å, 6900 Å, and 6290 Å and the Herzberg bands found within the blue between 3300 Å and 4600 Å (KRASSOVSKY et al., 1962). NOXON (1982) described the O_2 ($^1\Delta_g$) emission behavior for twilight periods.

Found in great proliferation across the visible airglow spectrum shown in part in the spectrum in Figure 2 (taken from the airglow atlas of BROADFOOT and KENDALL, 1968) are the emissions from the vibrational states of $OH(^2\Pi)$ cascading down from the initial state excited in the reaction of atomic hydrogen with ozone (BATES and NICOLET, 1950; BATES and MOISEIWITSCH, 1956). The molecular lines from these emissions for the most extended rotational branch, the P branch, are found to extend over many tens of Angstroms. Figure 3 shows an example of field-widened Michelson interferometer observations of the infrared OH airglow (DESPAIN et al., 1971; BAKER et al., 1981), where the components of the OH molecular band structure may be seen with high resolution for several bands located in the near infrared. A major breakthrough in the laboratory work in the last three years has provided detailed rotational analysis of the OH molecular bands. Consequently, synthetic spectra may be constructed with much improved spectroscopic constants for all bands up to v = 10 (GOLDMAN, 1982; COXON, 1980; COXON and FOSTER, 1982).
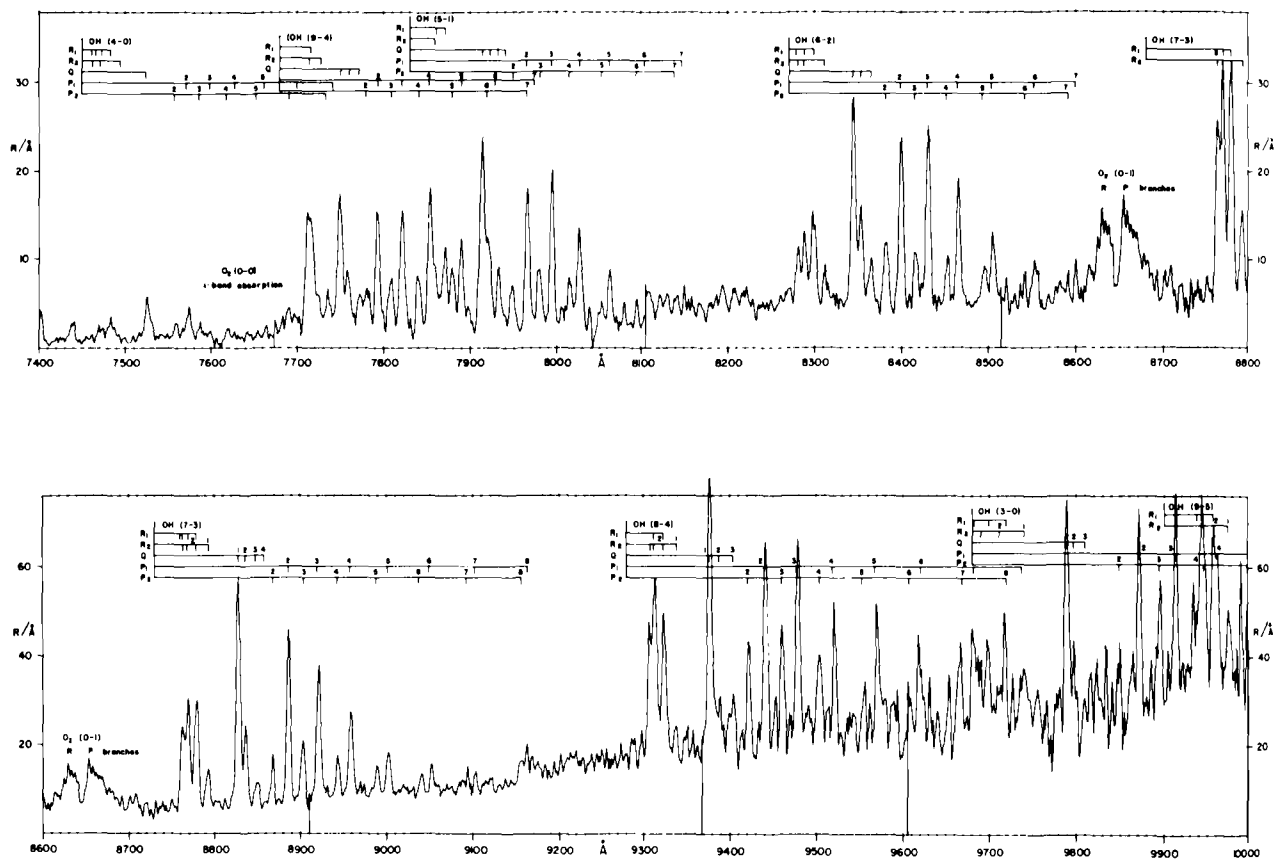
Figure 2. Selection of visible airglow spectrum showing OH emissions together
with other terrestrial nighttime features (from BROADFOOT and KENDALL, 1968).
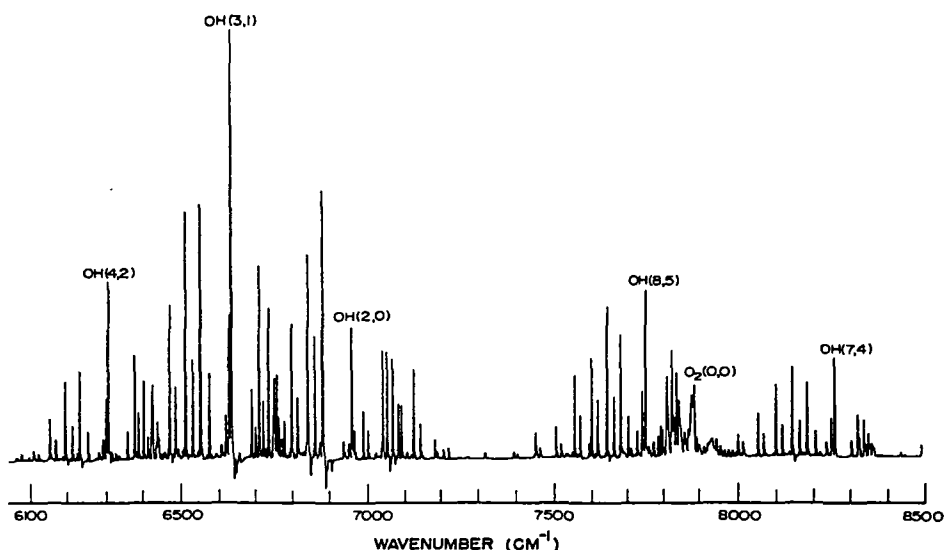
Figure 3. Selection of infrared airglow spectrum showing OH
emissions between 1 and 2 μ (from WARE, 1980).

Figure 4 shows an OH interferogram obtained in 1972 by the Michigan Airglow
Observatory with a Fabry-Perot interferometer (MERIWETHER et al., 1974). The
separation between the two OH peaks in each order corresponds to the lambda
splitting of the ground molecular state of OH induced by the spin orbit coupling
of the lambda states (HERZBERG, 1950). The difference here corresponds to
0.13 cm^{-1}, which is in good agreement with the result found by COXON and
FOSTER (1982). The doublet structure of the OH rotational line in high
resolution is a fact not widely appreciated. This work has been extended by
HERNANDEZ and SMITH (1984) who also demonstrated for the first time the
capability to measure directly mesospheric winds from the Doppler shift of OH
emissions measured by a Fabry-Perot interferometer.

SELECTION OF MOLECULAR EMISSIONS FOR ROTATIONAL TEMPERATURE MEASUREMENTS

The two most common determinations of the rotational temperature of meso-
sphere molecular emissions documented in the literature involve relative
intensity measurements or components of the O_2 atmospheric band and OH
molecular emissions. We note that the centroid height for any of the molecular
oxygen nightglow emissions is about 95 km; there is evidence that there may be
small differences of the order of 1 or 2 km between these layers (WITT et al.,
1979; LLEWELLYN and LONG, 1980).

The band emissions of the Herzberg I system are much weaker than the near
infrared emissions of the Atmospheric bands (DEGEN, 1972, 1977); therefore,
there is little advantage to be gained from the application of the Herzberg I
emissions for rotational temperature determinations. The results of early
observations and the theory for the synthetic spectrum for the Herzberg I
emissions are given by KRASSOVSKY et al. (1962).

The two band systems, the OH and the O_2 atmospheric band system, provide
a means by which the temperatures of two mesosphere heights may be monitored
simultaneously. The observations by NOXON (1978) showed that correlations can
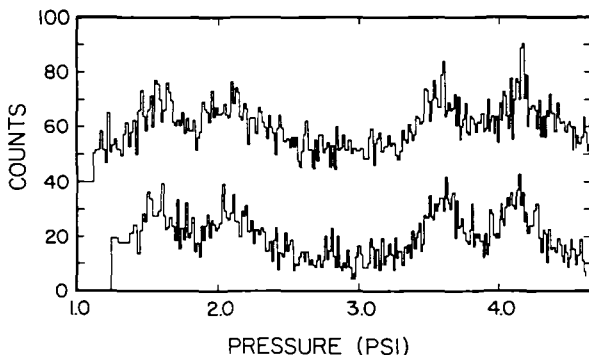
8/9 FEBRUARY 1972
OH FPI SPECTRA

Figure 4. High resolution Fabry-Perot spectrum showing
the doublet structure of the OH $P_1(2)$ rotational
line at 7316.2A'%. The spacer gap used was 1 cm.

be found in data sets involving simultaneous observations of both emissions; the analysis in his paper showed that upward progression of internal gravity waves was implied from the observed correlations. WEINSTOCK (1978) has used these results to examine theoretically the connection between observed effects in the surface brightness distribution of either OH or $O_2(^1\Sigma)$ emissions and the range of gravity wave spectra that may be observed through airglow intensity or temperature fluctuations.

The other molecular oxygen emissions found in mesospheric emissions are too weak for useful measurements of the rotational temperature or lie in a region of the spectrum where absorption by water vapor in the atmosphere is severe (such as the 1.27µ 0-1 band of the O_2 ($^1\Delta_g$) band system). The metallic emissions of species such as iron, magnesium, and calcium are not useful for studies of the mesosphere temperature.

THEORY OF ROTATIONAL TEMPERATURE ANALYSIS

The general formulation of the relation between the volume emission rate, I, of individual rotational lines and the rotational temperature, T_r, for a diatomic molecular band may be expressed (MIES, 1974; WARE, 1980) as

$$I(J'', v'' \leftarrow J', v') = N_{v'J'}(T_r)A(J'', v'' \leftarrow J', v') \qquad (1)$$

where $N_{v'J'}$ is the population of molecules in the initial state defined by the rotational and vibrational quantum numbers, J' and v', before transition to the state J'', v''. The coefficient $A_{v'J',v''J''}$ is the Einstein transition probability of spontaneous emission. In the case of thermal equilibrium, the population distribution of the rotational states for a given initial vibrational level of OH or O_2 may be described by the Maxwell-Boltzmann distribution. Therefore, the dependence of the volume emission rate (photons/cm^3/sec) of each rotational line observed on the rotational temperature may be written as

$$I_{v'J',v''J''} = k_{v'}(2J' + 1) A(J'', v'' \leftarrow J', v') X \qquad (2)$$
$$\exp[-F_v(J')hc/kT_r]$$

where the factor $(2J' + 1)$ is the statistical weight of the state J', and

$F_{v'}(J')$ is the term difference between the initial and final states. The constant, $k_{v'}$, is the reciprocal ratio of the rotational partition function and needs not be evaluated for a measurement of the rotational temperature. The ground-based instrument measures from the ground the column emission rate which is the quantity $I_{v'J'}$, integrated over the airglow layer as observed along the instrumental line of sight. Calculations made by SIVJEE et al. (1972) and MERIWETHER (1975) show that the mesospheric thermal variation across the region of the OH airglow was sufficiently small that the Boltzmann factor of equation (2) may be removed from the integrand. The same argument may be applied to the molecular oxygen emissions emitted in the lower thermosphere above the OH airglow layer with somewhat less confidence. Hence, equation (2) may be rewritten to arrive at an expression for the rotational temperature T_r:

$$T_r = \frac{hc/k[F_{v'}(J_1') - F_{v'}(J_2')]}{\ln(\frac{R_1(2J_2' + 1)A_2}{R_2(2J_1' + 1)A_1})} \tag{3}$$

where the factor hc/k is evaluated as 1.439 °K cm. $R_{J'v'}$ is the integrated column emission rate of the selected rotational line expressed in rayleighs (HUNTEN, 1956).

For a spectral scanning instrument observing several rotational lines, a plot of the function $F_{v'}(J')hc/k$ against the abscissa, $\ln(R_{v'J'}/(2J' + 1) A_{v'J'})$, will give the rotational temperature from the evaluation of the slope of the linear fit through the points. In the case described by MERIWETHER (1975; 1979) and TEPLEY et al. (1981), only the $P_1(2)$ and $P_1(5)$ lines of the 8-3 band with wavelengths 7316.2 Å and 7401.6 Å were used. In this application the expression for the rotational temperature is given simply as

$$T_r = 362/\ln[3.43\frac{R_2}{R_5}] \tag{4}$$

There are two salient points to be discussed in connection with the application of equation (3). First, the transition probabilities used to compute the rotational temperature in this expression are subject to error depending upon the precision of the quantum mechanical calculations entering into these results. In the case of OH, the values calculated by KVIFTE (1961) have been used by many investigators (SIVJEE et al., 1972; DICK, 1972; MYRABO et al., 1983). However, the more precise quantum-mechanical calculations carried out by MIES (1974) lead to vibrational Einstein coefficients that are in agreement with the observed intensities of the vibronic overtone emissions of OH that appear in the infrared between 1 to 3 microns (MURPHY, 1971). The Einstein coefficients for the rotational lines are not available in the Mies reference, but can be extracted from Mies's dissertation. These values are also listed in the doctoral dissertation by WARE (1980). MERIWETHER (1975) found the difference in the computation of the OH rotational temperature with Mies's set as compared with the older set to be a decrease of 3 to 6 percent. MYRABO et al. (1984) has pointed out, however, that the recent theoretical work by WERNER et al. (1983) show areas of disagreement in comparison with Mies's results. The maximum absolute error in OH measurements is less than a few degrees.

The spectroscopic constants and line strength values for O_2 are given by SCHLAPP (1936). An M.S. dissertation by SHOU-CHI SUE (1976) provides computer codes for the production of synthetic spectra and the determination of O_2 rotational temperature by curve fitting.

The other point to be discussed in connection with equation (3) is the selection of the P branch rotational lines to be used in the determination of the rotational temperature. Most investigators that have used scanning

spectrometers in this problem choose to scan the entire extent of the P branch.
MERIWETHER (1975, 1979) pointed out that the ratio of signal to noise may be
improved by greater sampling of selected rotational lines by programming the
scan to switch from a line such as the $P_1(2)$ line to another line, such as the
$P_1(5)$ line, while skipping the intermediate P branch lines. The justification
for this strategy is simply that the sensitivity of the intermediate lines to
changes of the rotational temperature is considerably less than the lines
located at the head or tail of the branch. There is no point in expending
instrumental time in measuring the intensity of a rotational line that does not
respond to variations of the rotational temperature. The OH spectra obtained by
MERIWETHER (1979) with a 1 m spectrophotometer required 5 minutes of integration
(including the scanning overhead) for 5 °K precision while the spectra obtained
by DICK (1972, 1977) with the same instrument required 15 minutes integration
for 15 °K precision. The price that must be paid, however, is that the back-
ground positions must be selected carefully, and the instrumental wavelength
drifts of the step position settings must be negligible over many hours of
observations.

SELECTION OF OPTICAL INSTRUMENTATION FOR ROTATIONAL TEMPERATURE OBSERVATIONS

The weak levels of OH emissions within the window of adequate sensitivity
for the photomultiplier, have required low resolution instrumentation with high
throughput to be mandatory for rotational temperature determinations.
Typically, photometers using narrow band interference filters possess the
greatest instrumental sensitivity but more care is required to achieve the level
of absolute accuracy that may be obtained by the grating instruments.

The first detection and identification of OH molecular emissions in the
nightglow were made by MEINEL (1950a,b,c) with a spectrograph where many hours
of exposure were necessary to obtain a reasonable precision. This avenue of
research has been continued by Krassovsky and his collaborators with image
intensifier optics attached to the spectrograph (KRASSOVSKY and SHAGAEV, 1974;
SHAGAEV, 1974; KRASSOVSKY et al., 1977). The superior properties of the
scanning spectrometer made the spectrograph design obsolete until recent years
when array detector technology provided detectors with much greater quantum
efficiency than the photographic plate. The field-widened Michelson inter-
ferometer (DESPAIN et al., 1971; BAKER et al., 1981) has been the most sensitive
instrument applied to this problem with results of high precision and excellent
temporal resolution. However, this instrument is complex and more expensive
than the others. Moreover, close attention to the instrumental alignment is
needed throughout the observations, so it is not well suited for automatic
observations.

In this section we summarize the properties of the two most commonly used
instruments that have been applied for observations of the rotational
temperature, the grating spectrometer and the interference filter photometer.

(a) Instrumental Theory of the Grating Spectrometer

The best design of the grating spectrometer commonly employed is the
Fastie-Ebert (FASTIE, 1952; 1962; 1967). This instrument employs a grating that
is tilted through the range of incidence angles that corresponds to the spectral
region of interest. Figure 5 shows a schematic layout for this instrument taken
from FASTIE (1967). The key to the understanding of this instrument is that the
spherical aberation introduced into the image upon the first reflection from the
Ebert mirror is largely canceled out by the second reflection from the Ebert
mirror after the ray has encountered the dispersing element of the grating.
Hence, the actual resolution of the instrument will approach the theoretical
resolution limit better than any other design involving spherical mirrors. The
remaining distortion of the slit image, which is astigmatism, may be dealt with
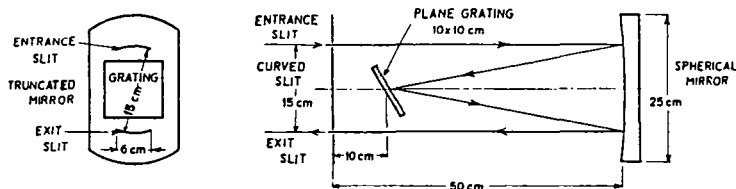
Figure 5. Schematic showing the optical configuration
for the Fastie-Ebert spectrometer design (from FASTIE,
1967).

by the innovation of curved slits that was introduced by FASTIE (1952).

The equation that relates the number of photons brought to the detector,
Φ, from the extended source region of surface brightness B viewed by the
instrument may be written (following the presentation of VALLANCE-JONES, 1974)
as

$$LR = (\Phi/B)R = A_g[\sin i + \sin \Theta]\beta\tau \tag{5}$$

where L is the luminosity of the extended source. R is the resolution defined
as $\lambda/\Delta\lambda$ specified by the physical slit width chosen, A_g is the grating area,
i the incidence angle, Θ the diffraction angle, τ the transmission of the
instrument including the grating efficiency and the transmission of blocking
filters employed, and β is the angular length of the slits.

As discussed by Vallance-Jones in his book, the luminosity equation shows
that the Ebert-Fastie design is at the point where further improvement in the
instrumental design itself will not result in major gains in instrumental
performance. The factor $[\sin i + \sin\Theta]$ is of order unity while the trans-
mission, τ, will hardly be greater than 50%; the angular slit length will
typically be about 15%. Hence, the sensitivity of the Ebert-Fastie grating
instrument may be enhanced only by enlarging the size of the instrument. This
would provide for a larger grating area but would require an Ebert mirror with a
longer focal length. The biggest grating in a ground-based instrument employed
for ground-based studies has been of size 15 cm x 25 cm for application in a 1 m
spectrometer (FASTIE, 1967; DICK, 1972; MERIWETHER, 1979). The next size
commonly used has been 10 cm x 12.5 cm for a 1/2 m spectrometer (GATTINGER and
VALLANCE-JONES, 1974). The plot in Figure 6 prepared by VALLANCE-JONES (1974)
shows the ratio of signal to noise for a set of typical values for the A_g, τ,
i, Θ parameters in equation (5) as a function of source brightness and dark
count level.

A second point is that the spectral image will be formed only at the
position of the exit slit. Hence, the only application of the Ebert-Fastie
design possible is in the scanning mode involving either the rotation of the
grating or the rotation of the Ebert mirror. The former is more normally
applied as the instrument is inherently more stable in this mode. Hence, any
application that seeks to employ position sensitive detectors such as the CCD
array used in imaging applications must use a spectrograph design.

The coupling of a photomultiplier with the scanning spectrometer is
superior in signal to noise performance as compared with the photographic
spectrograph. The quantum efficiency of the photomultiplier, which reaches
beyond 20% for the Ga-As RCA phototube, is about two orders of magnitude greater
than the photographic plate. This advantage is lost once the quantum efficiency
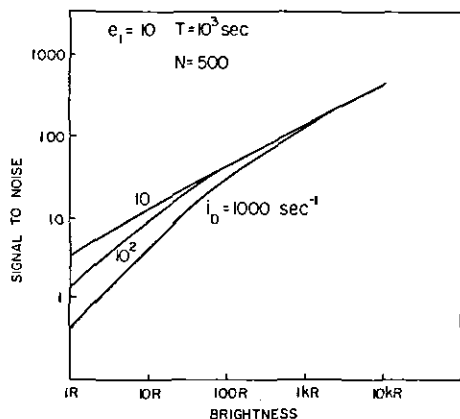
Figure 6. Plot of the ratio of signal to noise for several instrumental parameters of the spectrometer (from VALLANCE JONES, 1974).

of an individual pixel in the imaging spectrograph reaches that of the photo-multiplier (TORR et al., 1982). For most practical applications, the Ebert-Fastie spectrometer of sufficient throughput fitted with a good red sensitive photomultiplier will have suitable sensitivity for mesosphere temperature observations, but with moderate temporal resolution.

(b) Interference Filter Photometry

It is not appreciated that the interference filter is a low resolution solid state Fabry-Perot interferometer. The filter has two major components, one the part containing a multitude of dielectric layers with a composition either high or low index of optical thickness, eg. ZnS or cryolite. The design of the stack of layers is such that an etalon cavity simulating an etalon spacer of about 1 micron is formed. The blocker component of the filter serves as an edge filter that removes the part of the filter transmission spectrum (the stop band) at wavelengths several thousand Angstroms below the filter passband that otherwise would transmit the source emissions. The stop band at the red end of the transmission spectrum is removed by the lack of adequate sensitivity of the detector. A good filter that would serve the needs of OH rotational temperature observations would have two or more such cavities in series. The second cavity identical in design with the first cavity serves to sharpen the passband and to improve the rejection of the background continuum in the wings of the passband.

Figure 7 presents a schematic view of a design intended for the application of an interference filter of widths as narrow as 3 to 5 Å. Thermal control of the filter wheel housing is advisable as the thermal coefficient of band shifts for typical interference filters is of the order of 0.3 Å/°K.

The use of such narrow bandwidths for these filters improves the ratio of signal to noise by reducing the background transmitted by the filter so that operations near the phase of full moon or during early twilight are not greatly handicapped by the background levels. Moreover, small bandwidths make the selection of the background positions easier in those applications involving programmable scanning of the tilt cycle. Furthermore, the measurement of the background level is a cleaner measurement with the chances of contamination by weak molecular bands or atomic emission lines less severe. Such a filter is more expensive to purchase and its peak transmission will generally be smaller by factors from 30 to 50%.

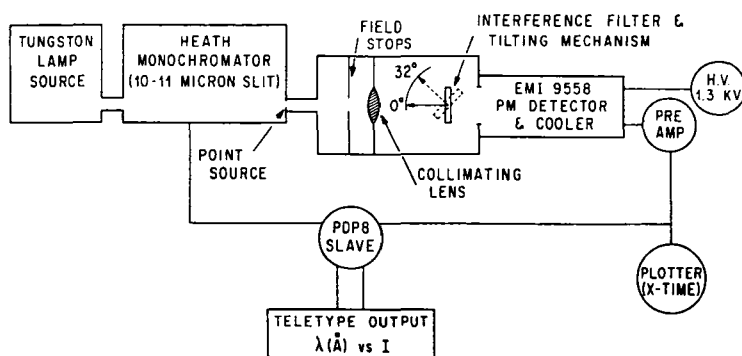Just as in the high resolution application of the Fabry-Perot

Figure 7.   Schematic showing the optical configuration
for the tilting filter photometer.

interferometer, scanning of the source spectrum may be accomplished by several
means.   The method generally used tilts the filter with respect to the optical
axis (EATHER and REASONER, 1969). The shift of the passband of the interference
filter is related to the tilt angle by the relation

$$\Delta\lambda = \alpha(\Delta\Theta)^2 \qquad (6)$$

where $\Delta\Theta$ is the tilt angle of the filter away from the normal, and $\Delta\lambda$ the shift
(in Angstroms) to the blue that takes place.   The proportionality constant, α,
is about 0.2 $\text{Å}/(\text{deg})^2$.

    The shape of the passband of the filter depends not only on the properties
of the materials used to fabricate the filter but also on the aperture function
defined by the circular aperture stop placed at the objective image plane. When
the filter is tilted, the aperture function changes as a function of the tilt
angle in a manner that maintains the total area of the instrumental function
constant.   Since the spatial distribution of the light from the interference
order generated by the filter decreases quadratically with angle, for the
aperture function to remain constant, the aperture must change its dimension in
the radial direction while maintaining constant area.   Since this is not
possible for a circular aperture, what happens is that the passband of the
filter broadens and the peak transmission decreases with tilt angle.   Figure 8
demonstrates this for several tilt angles.

    ROHRBAUGH et al. (1983) avoided in part this loss of transmission by
choosing an aperture with a crescent shape.   This innovation enabled the
measurement of the line intensity at a given tilt angle with maximum sensitivity
and the background continuum measured on both sides of the line with minor
degradation of the passband shape.

    SWENSON (1974) in his doctoral dissertion with P. B. Hays mechanically
modified the shape of the aperture while retaining constant area.   In this
application the actual aperture shape for a given tilt angle was defined by the
intersection of a rotating spiral aperture wheel and a wedge aperture. Figure 9
shows how the passband of the filter remains constant as a function of the
incident angle; this design avoids any transmission decrease resulting from
changes in the aperture function.   In this application the portion of the sky
viewed will shift as the aperture moves out from the optical axis.   When it is
just the background continuum that is being measured, this is not a serious
limitation.
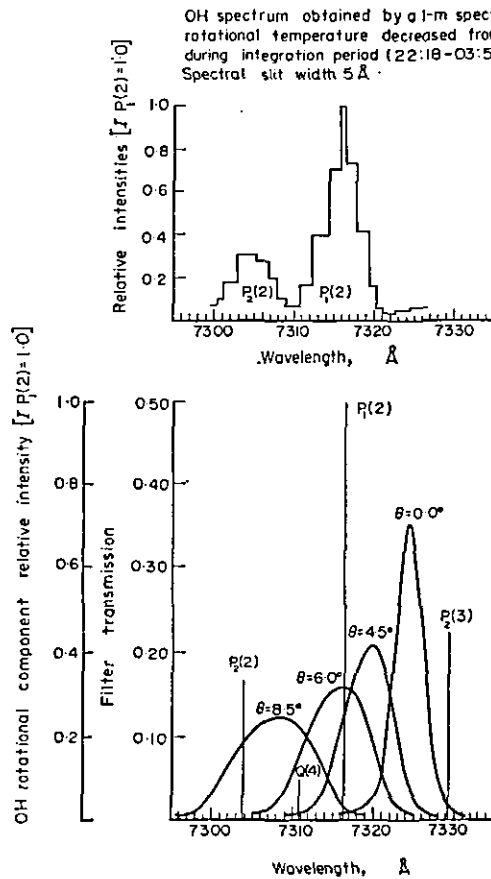
Figure 8. Calibration of the instrumental passband for
various scan positions for the Swenson-Hays scanning
photometer (from SWENSON, 1974).

The sensitivity of the photometer to an atomic line emission may be
written as

$$\Phi = \frac{10^6}{4\pi} A\Omega Q T_o T_f B_\lambda \qquad (7)$$

where A is the filter area, Ω the field of view of the instrument in
steradians, T_f the transmission of the filter at the line position, T_o the
transmission of the instrumental optics including mirror surfaces, lens,
windows, Q is the quantum efficiency of the photomultiplier, and B_λ is the
source surface brightness for the spectral line emission in question.
Typically, for a red sensitive photomultiplier with a field of view of 3° and
a filter of diameter of 5 cm, a sensitivity at 7300 Å of 25 counts/sec/R is
attained (MERIWETHER, 1975).

The first tilting filter photometer measurements of the OH rotational
temperature were obtained by VISCONTI et al. (1971). In this work, a single
filter was tilted through an angle sufficient to scan about 90 Å. Measurements
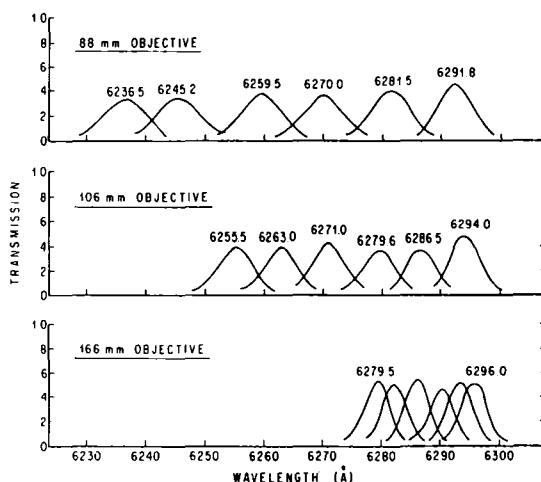
Figure 9. Comparison of the instrumental response for
two different scan modes of the tilting filter photo-
meter (from MERIWETHER, 1975).

were obtained at a rate of 1 averaged spectrum per hour with a precision of
about 10 percent. This work was extended by MERIWETHER (1975) who achieved a
precision of 5k for each measurement at a rate of 35 measurements per hour.

Several changes were implemented to achieve this improvement: the first was
the substitution of a EMI red sensitive photomultiplier in place of the less
sensitive one used by Visconti. The second was the application of two
interference filters that were switched into position in sequence. The greater
transmission of the filter closer to zero tilt angle enhances the throughput of
the instrument. The third difference was the increase of the sampling rate of
the rotational lines of OH achieved by the selection of just two tilt angles,
one designed to observe the peak of the rotational line and the other to measure
the background emission. This method was extended by TEPLEY et al. (1981) in
which a minicomputer interfaced with a stepping motor was used to carry out the
tilt scan of the filter in place of the original method involving the selection
of the desired tilt angle by mechanical means. Figure 10 depicts the
improvement in the sampling rate gained by the method of programmable selection
of the spectral position used.

The tilting filter photometer is not the only means for the measurement of
the OH rotational temperature by application of photometry. A filter wheel
photometer with a set of filters selected to measure the appropriate background
and the line emissions would be an alternative approach to the problem. The
advantage of the tilting filter photometer is that the same filter is used to
observe the background as for the line emission. The need to consider the
properties of different filters in their response to line and continuum
emissions, which would require extensive wavelength and intensity calibrations,
is thereby avoided.

CONCLUDING REMARKS

The question as to which of the various instrumental methods described here
or alluded to would be the method of choice for measuring rotational
temperatures of mesosphere airglow emissions cannot be answered directly.
Much depends upon the particular objective for the desired measurements in

High latitude airglow observations on hydroxyl 8-3 band

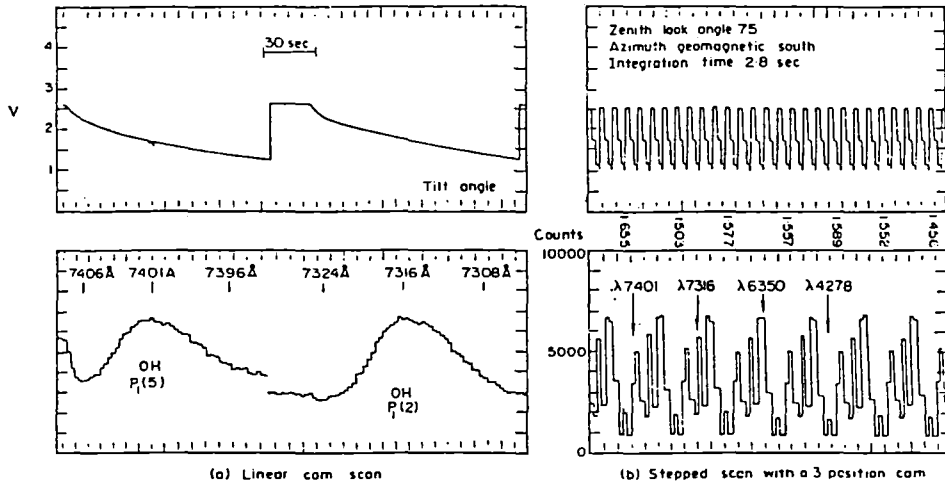(a) Linear cam scan   (b) Stepped scan with a 3 position cam

Figure 10. Example of data acquisition with the stepper cam. The strip
chart speed is identical for both a and b; the temperature computed
from the ratio of intensities of the two lines in a is 210 ± 5K.

addition to the particular resources available in terms of finances and
instruments.

If the purpose were to study the distribution of gravity waves in the
mesosphere thermal structure, then photometry is the only method of choice.
A grating spectrometer that matches the capability of the instrument employed by
MERIWETHER (1979) or MYRABO et al. (1983) would be almost an order of magnitude
more expensive than the photometry instrumentation. Of course, such a grating
spectrometer possesses much more versatility than the photometer and would be
more suitable as an airglow or auroral observatory instrument than the photo-
meter. Further refinement such as position sensitive detectors would only
increase the cost of the grating instrument and would require extensive software
development. The photometric instrumentation approach applying interference
filters possesses greater temporal resolution by at least an order of magnitude.
It is therefore possible to utilize various mapping strategies where tradeoffs
of the accuracy of the measurement in a particular direction are made for the
increased number of directions taken in for the map. However, the photometer
requires more care in application for accurate absolute measurements. However,
if the aim of the measurement was to determine the mesosphere temperature with
temporal resolution of a few minutes, then a grating spectrometer will perform
this task reasonably well. However, this problem is pretty much restricted to
the use of the OH emission lines for mapping applications because the O_2
atmospheric band at 8650 Å is not so amenable to photometric application as the
OH rotational lines.

Should the problem be more one where good absolute measurements of the
rotational temperature is desired in conjunction with other experiments, then
the grating spectrometer becomes more the instrument of choice. The analysis of
the observations is more straightforward and the concurrent requirement for
instrumental calibration not nearly so stringent for the spectrometer instrument
as it is for the photometer. The best mode of study would be the combination of
the two instruments where the grating spectrometer serves to calibrate the

15

photometer observations. Then it would be possible to acquire accurate
observations with good temporal resolution.

It is assumed that any application of either instrument would take
advantage of the minicomputer in its design. Such devices are so inexpensive
these days and the time and effort saved in data acquisition and analysis would
pay for the expense of this hardware in such short order that it would seem
obvious that this decision would be the most fruitful in scientific productivity
(MERIWETHER, 1979).

While the focus on the experimental objective has been placed on the
science that may be attained with the measurements alone, collaborative
observations with other means of studying the mesosphere would have numerous
useful results. As an example, TV imaging systems involving the application of
broadband filters with image orthicon have shown an ability to observe the sky
brightness distribution of the OH emission in real time. One can imagine
experiments where the photometric instrument is applied to observe the
rotational temperature distribution across brightness irregularities detected in
real time by the TV system. The combination of the two approaches would serve
to characterize the gravity wave feature more specifically than possible with
either instrument alone. Other instruments relating to the study of the meso-
sphere that would profit from collaborative studies with optical instrumentation
would be the MST radar and meteor radar that observe horizontal and vertical
winds in the mesosphere and lower thermosphere.

REFERENCES

Arnold, F. (1980), Ion-induced nucleation of atmospheric water vapor at the
    mesopause, Planet. Space Sci., 28, 1003-1009.
Arnold, F., D. Krankowsky, E. Zettwitz and W. Joos (1980), Strong temperature
    control of the ionospheric D-region evidence from in-situ ion composition
    measurements, J. Atmos. Terr. Phys., 42, 249.
Baker, D., A. Steed and A. T. Stair, Jr. (1981), Development of infrared
    interferometry for upper atmospheric emission studies, Appl. Optics, 20,
    1734-1746.
Bates, D. R. (1981), The green light of the night sky, Planet. Space Sci.,
    29, 1061-1067.
Bates, D. R. (1982), Airglow and auroras, App. Atomic Collision Phys., 1,
    149-224.
Bates, D. R. and B. L. Moiseiwitsch (1956), Origin of the Meinel hydroxyl
    system in the night airglow, J. Atmos. Terr. Phys., 8, 305-308.
Bates, D. R. and M. Nicolet (1950), The photochemistry of atmospheric water
    vapor, J. Geophys. Res., 55, 301-327.
Beynon, W. J. G. and E. S. O. Jones (1965), 'Winter Anomaly' in ionospheric
    absorption and stratospheric warmings, Nature, 206, 1242-1245.
Bjorn, L. G. and F. Arnold (1981), Mass spectrometric detection of
    precondensation nucei at the arctic summer mesopause, Geophys. Res. Letts.,
    8, 1167-1170.
Broadfoot, A. L. and K. R. Kendall (1968), The airglow spectrum, 3100-10,000 Å,
    J. Geophys. Res., 73, 426-428.
Chakrabarty, D. K. and P. Chakrabarty (1978), An attempt to identify the
    obscured paths of water cluster ions build-up in the D region, J. Atmos.
    Terr. Phys., 40, 437.
Chakrabarty, P. and D. K. Chakrabarty (1979), Mesopause temperature from the
    $[NO^+ H_2O]/[NO^+]$ ratio, J. Geophys. Res., 84, 3403.
Cogger, L. L. and J. S. Murphree (1980), The latitudinal and seasonal variation
    of atomic oxygen deduced from observations of the E-region OI 557.7 nm
    airglow, Space Res., 20, 115-120.
Cogger, L. L., R. D. Elphinstone and J. S. Murphree (1981), Temporal and
    latitudinal 5577 Å airglow variations, Can. J. Phys., 59, 1296-1307.

16

Coxon, J. A. (1980), Optimum molecular constants and term values for the
X(v $\leq$ 5) and A(v $\leq$ 3) state of OH, <u>Can. J. Phys.</u>, <u>58</u>, 933.

Coxon, J. A. and S. C. Foster (1982), Rotational analysis of hydroxyl vibration-
rotation emission bands: Molecular constants for OH X,6 $>$ v $\leq$ 10, <u>Can. J.
Phys.</u>, <u>60</u>, 41.

Degen, V. (1972), Excitation of the Herzberg bands of O_2 in laboratory after-
glow and night airglow, <u>J. Geophys. Res.</u>, <u>77</u>, 6213-6218.

Degen, V. (1977), Nightglow emission rates in the O_2 Herzberg bands, <u>J.
Geophys. Res.</u>, <u>82</u>, 2437-2438.

Despain, A. M., D. J. Baker, A. J. Steed and T. Tohmatsu (1971), Extended-field
large-aperture interferometer-spectrometer for airglow surveys, <u>Appl.
Optics</u>, <u>10</u>, 1870-1876.

Dick, K. A. (1972), Short time-interval spectrometric hydroxyl emission studies,
<u>Ann. Geophys.</u>, <u>28</u>, 149-153.

Dick, K. A. (1977), On the rotational temperature of airglow hydroxyl emissions,
<u>Planet. Space Sci.</u>, <u>25</u>, 595-596.

Donahue, T. M., B. Guenther and J. E. Blamont (1972), Noctilucent clouds in
daytime: circumpolar particulate layers near the summer mesopause, <u>J.
Atmos. Sci.</u>, <u>29</u>, 1205-1209.

Donahue, T. M., B. Guenther and R. J. Thomas (1974), Spatial and temporal
behavior of atomic oxygen determined by Ogo 6 airglow observations, <u>J.
Geophys. Res.</u>, <u>79</u>, 1959-1964.

Eather, T. H. and D. L. Reasoner (1969), Spectrophotometry of faint light
sources with a tilting-filter photometer, <u>Appl. Optics</u>, <u>8</u>, 227-242.

Fastie, W. G. (1952), A small plane grating monochromator, <u>J. Opt. Soc.</u>,
<u>42</u>, 641-651.

Fastie, W. G. (1967), Ultraviolet measurements in planetary atmosphere,
<u>Appl. Optics</u>, <u>6</u>, 397-402.

Fastie, W. G., H. M. Crosswhite and T. P. Markham (1961), Far ultra-violet
auroral spectra with a rocket Ebert spectrophotometer, <u>Ann Geophys.</u>, <u>17</u>,
109-114.

Forbes, J. M. (1982), Temperature and solar zenith angle control of D-region
positive ion chemistry, <u>Planet. Space Sci.</u>, <u>30</u>, 1065-1072.

Gadsden, M. (1969), Antarctic twilight observations 1. Search for metallic
emission lines, <u>Ann. Geophys.</u>, <u>25</u>, 667-677.

Gadsden, M. (1970), Metallic atoms and ions in the upper atmosphere, <u>Ann.
Geophys.</u>, <u>26</u>, 141-150.

Gadsden, M. (1981), The silver-blue cloudlets again: Nucleation and growth of
ice in the mesosphere, <u>Planet. Space Sci.</u>, <u>29</u>, 1079-1087.

Gattinger, R. L. and A. V. Jones (1972), Observation and interpretation of
hydroxyl airglow emissions, <u>Physics and Chemistry of Upper Atmospheres</u>,
184-192, B. M. McCormac (ed.), D. Reidel, Dordrecht-Holland.

Gattinger, R. L. and A. V. Jones (1974), Quantitative spectroscopy of the aurora
II the spectrum of medium intensity aurora between 4500 and 8900 Å, <u>Can. J.
Phys.</u>, <u>52</u>, 2343.

Goldman, A. (1982), Line parameters for the atmospheric band system of OH,
<u>Appl. Optics</u>, <u>21</u>, 2100-2102.

Haurwitz, B. and B. Fogle (1969), Wave forms in noctilucent clouds, <u>Rev. Space
Phys.</u>, <u>16</u>, 85-95.

Herzberg, G. (1950), <u>Spectra of Diatomic Molecules</u>, D. van Nostrand Company,
Inc., Princeton, NJ.

Hernandez, G. and J. L. Smith (1984), Mesospheric wind determinations and the
$P_1(2)_{c,d}$ lines of the $X^2\Pi$ OH (8-3) band, <u>Geophys. Res. Letts.</u>, <u>11</u>,
534-537.

Ismail, S. and L. L. Cogger (1982), Temporal variations of polar cap OI 5577 Å
airglow, <u>Planet. Space Sci.</u>, <u>30</u>, 865-873.

Kieffaber, L. M. and A. W. Peterson (1973), Structure and fluctuations in the
OH airglow at 1.65 µ, <u>J. Atmos. Terr. Phys.</u>, <u>35</u>, 2013-2022.

Killeen, T. L. and P. B. Hays (1981), Nonthermal line profiles from $O(^1S)$ in
the thermospheric nightglow, <u>Appl. Optics</u>, <u>20</u>, 3314-3317.

Krassovsky, V. I., B. P. Potapov, A. I. Semenov, V. G. Sobolev, M. V. Shagaev and N. N. Shefov (1977), On the equilibrium nature of the rotational temperature of hydroxyl airglow, Planet. Space Sci., 25, 596-597.

Krassovsky, V. I. and M. V. Shagaev (1974), Optical method of recording acoustic or gravity waves in the upper atmosphere, J. Atmos. Terr. Phys., 36, 373-375.

Krassovsky, V. I. and N. N. Shefov (1962), Atlas of the airglow spectrum 3000-12400 Å, Planet. Space Sci., 9, 883-915.

Kvifte, G. (1961), Temperature measurements from OH bands, Planet. Space Sci., 5, 153-157.

Llewellyn, E. J. and B. H. Long (1978), The OH Meinel bands in the airglow - the radiative lifetime, Can. J. Phys., 56, 581-586.

Llewellyn, E. J., B. H. Solheim, G. Witt, J. Stegman and R. G. H. Greer (1980), On the excitation of oxygen emissions in the airglow of the terrestrial planets, J. Phys., 12, 179-183.

Meinel, A. B. (1950a), OH emission bands in the spectrum of the night sky. I., Astrophys. J., 111, 555-564.

Meinel, A. B. (1950b), OH emission bands in the spectrum of the night sky. II., Astrophys. J., 112, 120-130.

Meinel, A. B. (1950c), Hydride emission bands in the spectrum of the night sky, Astrophys. J., 111, 207-218.

Meriwether, J. W., Jr. (1975), High latitude airglow observations of correlated short-term fluctuations in the hydroxyl Meinel 8-3 band intensity and rotational temperature, Planet. Space Sci., 23, 1211-1221.

Meriwether, J. W., Jr. (1979), Measurements of weak airglow emissions with a programmable scanning spectrophotometer, Planet. Space Sci., 27, 1221-1232.

Meriwether, J. W., Jr., P. B. Hays, K. D. McWatters and A. F. Nagy (1974), Interferometric measurements of the 7319 A doublet emissions of OII, Planet. Space Sci., 22, 636-638.

Mies, F. H. (1974), Calculated vibrational transition probabilities of $OH(X^2\Pi)$, J. Mol. Spect., 53, 150-188.

Moreels, G. and M. Herse (1977), Photographic evidence of waves around the 85 km level, Planet. Space Sci., 25, 265-274.

Murphy, R. E. (1971), Infrared emission of OH in the fundamental and first overtone bands, J. Chem. Phys., 54, 4852-4859.

Myrabo, H. K., C. S. Deehr and G. G. Sivjee (1983), Large-amplitude nightglow OH (8-3) band intensity and rotational temperature variations during a 24-hour period at 78N, J. Geophys. Res., 88, 9255-9259.

Myrabo, H. K., G. J. Romick, G. G. Sivjee and C. S. Deehr (1984), Night airglow OH (8-3) band rotational temperatures at Poker Flat, Alaska, J. Geophys. Res., 89, 9153-9156.

Noxon, J. F. (1970), Optical emission from $O(^1D)$ and $O_2(^1\Sigma)$ in ultra-violet photolysis of O_2 and CO_2, J. Chem. Phys., 52, 1852.

Noxon, J. F. (1978), Effect of internal gravity waves upon night airglow temperatures, Geophys. Res. Letts., 5, 25-27.

Noxon, J. F. (1982), A global study of O_2 airglow: day and twilight, Planet. Space Sci., 30, 545-557.

Offermann, D., H. G. Bruckelmann, K. M. Torkar and H. Widdel (1981), Mesospheric structure and the D-region winter anomaly, Adv. Space Res., 1, 123-126.

Petitdidier, M. and H. Teitelbaum (1977), Lower thermosphere emissions and tides, Planet. Space Sci., 25, 711-721.

Rohrbaugh, R. P., B. A. Tinsley, H. Rassoul, Y. Sahai, N. R. Teixeira, R. G. Tull, D. R. Doss, A. L. Cochran, W. D. Cochran and E. S. Barker (1983), Observations of optical emissions from precipitation of energetic neutral atoms and ions from the ring current, J. Geophys. Res., 88, 6317-6330.

Schlapp, R. (1936), Fine structure in the $^1\Sigma$ ground state of the oxygen molecule, and the rotational intensity distribution in the atmospheric oxygen band, Phys. Revs., 51, 342.

Shagaev, M. V. (1974), Fast variations of hydroxyl night airglow emission, J. Atmos. Terr. Phys., 36, 367-371.

Sivjee, G. G., K. A. Dick and P. D. Feldman (1972), Temporal variations in night-time hydroxyl rotational temperature, Planet. Space Sci., 20, 261-269.

Solomon, S., G. C. Reid, R. G. Roble and P. J. Crutzen (1982), Photochemical coupling between the thermosphere and the lower atmosphere 2. D region ion chemistry and the winter anomaly, J. Geophys. Res., 87, 7221-7227.

Sue Shou-Chi (1976), Computer-aided estimates of the rotational temperatures of O_2 in the mesosphere, M.S. dissertation, Utah State University, Logan, Utah.

Swenson, G. R. (1974), A study of a sector spectrophotometer and auroral $O^+(^2P-^2D)$ emissions, Ph.D. dissertation, University of Michigan, Ann Arbor, Michigan.

Tepley, C. A., R. G. Burnside and J. W. Meriwether, Jr. (1981), Horizontal thermal structure of the mesosphere from observations of OH(8-3) band emissions, Planet. Space Sci., 29, 1241-1249.

Torr, M. R. and J. Devlin (1982), Intensified charge coupled devices for use as a spaceborne spectrographic image-plane detector system, Appl. Optics, 3091-3108.

Turco, R. P., O. B. Toon, R. C. Whitten, R. G. Keesee and D. Hollenback (1982), Noctilucent clouds: simulation studies of their genesis, properties and global influences, Planet. Space Sci., 30, 1147-1181.

Vallance Jones, A. (1974), Aurora, Geophysics and Astrophysics Monographs, D. Reidel, Dordrecht-Holland.

Visconti, G., F. Congeduti and G. Fiocco (1971), Fluctuations in the intensity and excitation temperature in the OH airglow (8-3) band, The Radiation Atmosphere, D. Reidel, Dordrecht-Holland, ed. B. M. McCormac, 82-89.

Ware, G. A. (1980), OH rotational temperatures using optimal interferometric techniques, Ph.D. dissertation, Utah State University, Logan, Utah.

Webb, W. L. (1965), Morphology of noctilucent clouds, J. Geophys. Res., 70, 4463-4465.

Weinstock, J. (1978), Theory of the interaction of gravity waves with O_2 airglow, J. Geophys. Res., 83, 5175-5185.

Werner, H.-J., P. Rosmus and E.-A. Reinsch (1983), Molecular properties from MCSCF-SCEP wave functions, 1, Accurate dipole moment functions of OH, OH^-, and OH^+, J. Chem. Phys., 79, 905-916.

Witt, G., J. Stegman, B. H. Solheim and E. F. Llewellyn (1979), A measurement of the O_2 atmospheric band and the $OI(^1S)$ green line in the nightglow, Planet. Space Sci., 27, 341-350.

2. APPLICATION OF FABRY-PEROT SPECTROMETERS FOR MEASUREMENT OF UPPER ATMOSPHERE TEMPERATURES AND WINDS

F. Jacka

The Mawson Institute for Antarctic Research
University of Adelaide

ABSTRACT

   The mechanisms giving rise to certain airglow and auroral emissions and their height distributions are briefly surveyed.

   The basic principles of operation of scanning Fabry-Perot spectrometers are outlined in order to illustrate the calibration and use of such instruments to measure Doppler broadening and shift of atomic emission lines in the spectrum of the airglow and aurora and hence to infer atmospheric temperatures and wind velocities. Technical details of some specific instruments are given. Brief mention is made of the use of wide angle Michelson interferometers for the same application. Emphasis is given to the future importance of imaging Fabry-Perot spectrometers to permit the detailed mapping of temperature and wind velocity, especially in the auroral regions.

   Some details are tabulated for those Fabry-Perot spectrometers known to be available for upper atmosphere temperature and wind measurement.

INTRODUCTION

   The use of a Fabry-Perot spectrometer (FPS) to infer upper atmosphere temperatures and winds is based on the measurement of airglow or auroral atomic emission line Doppler widths and displacements. The line width at half intensity is

$$\delta\lambda = 2(\ln 2)^{1/2} \, (2kT/M)^{1/2} \lambda/c = 7.16 \times 10^{-7} \lambda(T/M)^{1/2}$$

where λ is the centre wavelength of the emitted line, M the atomic mass and T the temperature. The line of sight component of the wind velocity is

$$V = c(\lambda' - \lambda)/\lambda$$

where λ' is the observed Doppler shifted wavelength. The assumption that T represents the bulk neutral atmosphere temperature implies equilibrium between the emitting species and the dominant neutral species. Whether this is an accurate assumption or not depends on the processes of excitation and emission. Whether this temperature can be assigned to a particular narrow height interval depends on the height distribution of the emitting species.

   The emissions to which this technique have been applied are the $\lambda630$ nm line from $O(^1D)$ at heights around 250 km and the $\lambda557.7$ nm line from $O(^1S)$ at heights around 97 km. In principle the technique should be applicable to the $\lambda589$ nm sodium line emitted around 90 km and perhaps to selected lines in the hydroxyl bands from about 85 km.

   By observing the Doppler shift of an emission line from a metastable ion such as the $\lambda732$ nm line from $O^+(^2P_{1/2})$ one may infer the line of sight component of the $\underline{ExB}/B^2$ drift and hence of $E^\perp$, the component of the electric field perpendicular to the geomagnetic field B.

EMISSION PROCESSES

The emission processes of greatest interest are briefly discussed in the following paragraphs. A more detailed account is presented by McEWAN and PHILLIPS (1975).

(a) $O(^1D)$ - 630.03 nm

The $\lambda 630$ nm emission from the thermosphere, around 250 km, has, in the past, been the principal subject of interest for measurement of upper atmosphere temperatures and winds using Fabry-Perot spectrometers. Consequently most of the literature in the field relates to it and most of the instruments have been designed primarily for measurements on this emission. For these reasons some comment on the emission processes are included even though the primary interests of this volume relate to lower levels.

The $O(^1D)$ at night is produced mainly by dissociative recombination of O_2^+ and electrons. This reaction also produces $O(^1S)$ which emits at $\lambda 557.7$ nm to produce $O(^1D)$.

An observed height distribution of $\lambda 630$ nm nightglow emission rate is illustrated in Figure 1 which is derived from the results of GULLEDGE et al., (1968).

The radiative lifetime of $O(^1D)$ is 110 sec. It is effectively quenched by N_2 with the consequence that the $\lambda 630$ nm emission is weak at heights below about 200 km. The mean time between collisions (~0.3 sec) in the region of the $\lambda 630$ nm emission is small compared with the $O(^1D)$ radiative lifetime so there is little doubt that the Doppler temperature is an accurate measure of the bulk kinetic temperature of the atmosphere.

During the day additional processes are operative viz. photoelectron impact excitation of ground state atomic oxygen, $O(^3P)$, and photodissociation of O_2 (WALLACE and McELROY, 1966); the emission has a broad peak around 150-250 km.

Typical $\lambda 630$ nm airglow intensities are ~100 R at night and up to 2.5 kR at noon.

In the aurora the dominant processes generating $O(^1D)$ are energetic particle impact on $O(^3P)$ and dissociative recombination of O_2^+ (produced by energetic particle impact on O_2). The resultant height distribution of the $\lambda 630$ nm emission is very broad and varies greatly with energy spectrum of the primary auroral particles. In general it cannot be estimated with useful accuracy from ground based observations. The distribution calculated by VALLANCE-JONES (1974) for a particular model primary particle spectrum shows half intensity values at ~120 and 320 km. Because of the 110 sec radiative lifetime of $O(^1D)$ the temperature inferred from the $\lambda 630$ nm line width will be a good measure of bulk neutral gas temperature at all levels. In the aurora the $\lambda 630$ nm intensity ranges up to ~100 kR.

(b) $O(^1S)$ - 557.73 nm

In the night airglow, as indicated above, some $O(^1S)$ is produced in the F region but the dominant processes are at about 97 km viz.

$$O + O + O \rightarrow O_2 + O(^1S) \qquad \text{(Chapman reaction)}$$

and

$$O + O + M \rightarrow O_2^* + M \qquad \text{(Barth reactions)}$$

$$O_2^* + O \rightarrow O_2 + O(^1S)$$
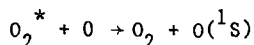
The relative importance of these reactions is discussed by BATES (1981). In both reaction schemes the rate of production of $O(^1S)$ is proportional to the third power of atomic oxygen density which peaks at about 97 km. At levels much below this quenching by O_2 and O compete with radiation at $\lambda558$ nm.

A measured height distribution showing the two emission regions is shown in Figure 1. More recent measurements in the lower region show height distributions peaked between ~92 - 100 km, about 7 km wide between half intensity levels and with quite rapidly varying structure attributed to gravity waves. An example is shown in Figure 2 from the work of THOMAS and YOUNG (1981).

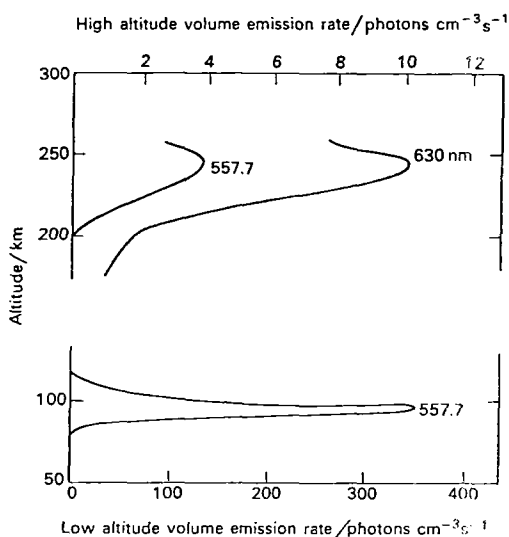Figure 1. Altitude profiles of the night airglow $\lambda558$ nm and $\lambda630$ nm emissions from atomic oxygen (from data of GULLEDGE et al., 1968).
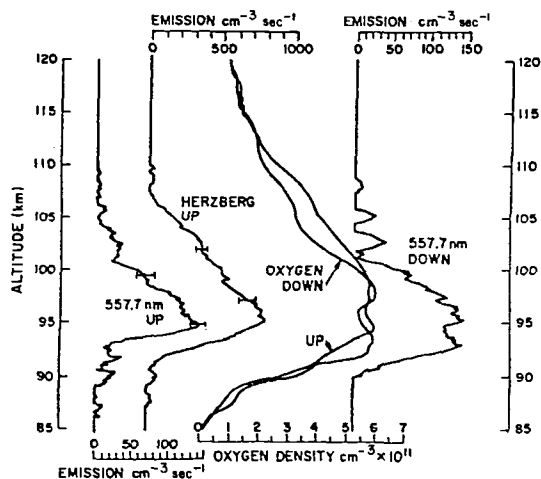

Figure 2. From left to right is shown the OI $\lambda557.7$ nm green line volume emission rate on the upleg, the Herzberg emission upleg, the oxygen density measured on both upleg and downleg, and the green line on the downleg (from THOMAS and YOUNG, 1981).

In the 97 km region the mean time between collisions is small compared with the radiative lifetime (0.74 sec) of $O(^1S)$ so these atoms are fully thermalized before emission. This is not so, however, at F region heights and because of the excess energy carried out by the $O(^1S)$ produced by dissociative recombination of O_2^+ account must be taken of this non-thermal contribution in interpreting the results of line profile measurements at $\lambda 558$ nm (HERNANDEZ, 1971; KILLEEN and HAYS, 1981).

Typically night airglow intensities of $\lambda 558$ nm are in the range 100-200 R, the F region contribution to this being about 20%.

Daytime emission of $\lambda 558$ nm has not been much studied. In addition to the nighttime processes, photoelectron excitation of atomic oxygen (WALLACE and McELROY, 1966) and photodissociation of O_2 (LAWRENCE and McEWAN, 1973) appear to be important contributors to $O(^1S)$ production. A measurement of $\lambda 558$ nm emission rate versus height is illustrated in Figure 3; only about 30% of the total emission comes from the lower region centered around 100 km.
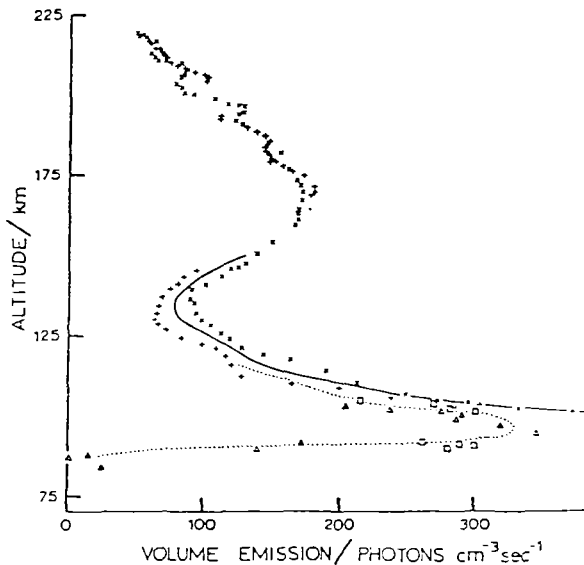
Figure 3. Observed OI 558 nm dayglow emission rate. The low-altitude emission, corrected for Rayleigh scattering, is indicated by squares and triangles (From WALLACE and McELROY, 1966).

Observed total intensities of 558 nm in daytime are of order 1 kR. As in the case of the nightglow, the contribution from levels above about 175 km may be significantly non-thermal.

In the aurora $\lambda 558$ nm intensities up to ~100 kR are a dominant visual feature. The $O(^1S)$ is excited mainly by electron impact and dissociative recombination of O_2^+. The intensity-height profile varies considerably with energy spectrum of the primary auroral particles. A model computation by VALLANCE-JONES (1974) yields half intensity levels at about 105 and 140 km. As in the dayglow, emission from heights above ~175 km may be significantly non-thermal.

(c) Sodium D Lines

The D lines of sodium are emitted from ~90 km height. At night the intensity of the D_2 $\lambda 589.0$ nm line is ~40 R. The excited sodium is probably produced mainly by the reaction

$$Na\ O + O \rightarrow Na\ (^2P) + O_2$$
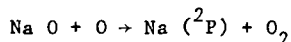
(KIRCHHOFF, 1983). Estimates of the noon and midnight concentrations of Na O versus height are shown in Figure 4.

Figure 4 also shows estimates of the noon and midnight concentrations of atomic sodium. Throughout twilight and day atomic sodium resonantly scatterers the D lines from sunlight; the intensity of the D_2 line reaches several kR. Because of the small noon—midnight change in Na distribution one may reasonably expect that the height and thickness of the $\lambda 589$ nm emission layer will vary little during the day.
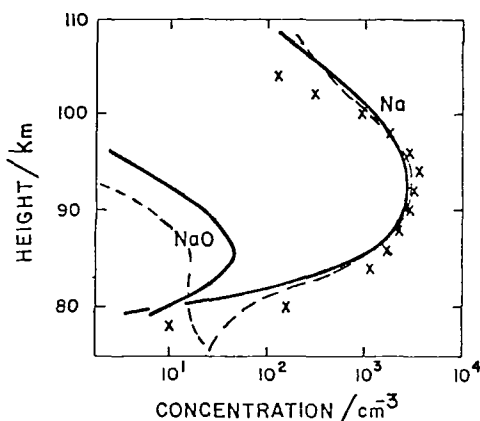
Figure 4. Sodium constituents calculated for noon (dashed curves) and midnight (full curves). The crosses represent an average of experimental profiles at midnight (from KIRCHHOFF, 1983).

Some twilight and night measurements on Doppler broadening of the D_2 line have been made by SIPLER and BIONDI (1978) and HERNANDEZ (1975). With good instrumentation it should be straightforward to measure Doppler broadening and shift of the D_2 line throughout the day. Such measurements are planned to commence here in 1984; none have been reported to date. Daytime observing will be aided by the fact that the solar spectrum shows a broad absorption feature at this wavelength, the intensity of the centre of the line being only 5% of the solar continuum. The dependence of line shape on temperature is discussed by CHAMBERLAIN (1961).

The sodium lines are not excited in the auroral spectrum.

(d) Hydroxyl Lines

Except in the infra-red the hydroxyl airglow bands are quite weak. However, it is not out of the question that with good instrumentation one might determine Doppler temperatures and winds from observations on say the $\lambda 727.7$ nm Q(1) line in the (8,3) band with typical intensity about 60 R emitted from the 80-90 km region.

Alternatively the OH rotational temperatures at this level may be deter-mined from measurements of relative intensities of lines in the (8,3) bands.

These may be recorded with a very low resolution scanning FPS with a bandwidth of say 0.5 nm, free spectral range ~15 nm and using a 3 period interference filter to select the required region of the spectrum. The basis for these suggestions is the work of BASEDOW (1977).

Other techniques for measurement of OH rotational temperatures are discussed by Meriwether in this volume.

(e) $O^+(^2P) - \lambda732$ nm

This is excited by photodissociation of O_2 during the day and by electron impact ionization of O in the aurora. The radiative lifetime of $O^+(^2P)$ is 4.6 sec. Quenching by N_2 is effective below about 200 km.

At 10° solar depression angle in twilight $\lambda732$ nm intensities of ~100 R were observed by MERIWETHER et al. (1978) using a grating spectrometer. These authors point out that such observations can be used to infer O and N_2 densities in the thermosphere. The same observations could be made with a FPS which will also permit the measurement of Doppler displacement and hence ion drift velocity and electric field – more precisely the line of sight component of the component of E perpendicular to B. Because of the high conductivity parallel to the magnetic field $E^\perp$ varies little down to the base of the ionosphere.

Auroral intensities at $\lambda732$ nm are usually very low, ~20 R, but at times of high fluxes of very low energy electrons, such as occur frequently in the polar cusp region, intensities of ~100 R are observed; Doppler displacements of ~2 pm, corresponding to ion drift velocities of ~800 ms^{-1}, are observed (SMITH et al., 1982).

FABRY-PEROT SPECTROMETER CHARACTERISTICS

For recording airglow or auroral emission line profiles the spectrometer usually incorporates a multi-layer dielectric interference filter to isolate the required spectral region and a high resolution, scanning Fabry-Perot etalon. For daytime observations an additional lower resolution etalon is required to permit adequate rejection of scattered sunlight near the wavelength of interest. The output is usually recorded with a photomultiplier or with a photoelectric imaging device.

The required bandwidth of the Fabry-Perot etalon is comparable with the width of the line to be recorded – typically a few picometres (pm). For measurement of wind velocity with a precision of 5 ms^{-1} the wavelength of the centre of this line must be determined with a precision of ~0.01 pm.

The basic theory of the ideal Fabry-Perot etalon is presented in standard optics texts e.g., BORN and WOLF (1965). The theory for non-ideal etalons has been developed by CHABBAL (1953), HERNANDEZ (1966, 1970, 1974) and others and is outlined here.

The phase difference between successive multiply reflected wavefronts emerging from the etalon is $2\pi m$ where m is the order of interference defined by

$$m \, \lambda/2 = \mu \, t_g \, \cos\theta + \lambda\phi/2\pi$$

μ is the refractive index of the spacer medium,
t_g is the geometrical spacing between the etalon plates,
θ is the angle of incidence of rays between the plates,
λ is the vacuum wavelength
ϕ a function of λ and θ, is the phase change on reflection at each of the plate surfaces.

For later convenience we write

$$m\lambda = 2(t_p + p) \cdot \cos\theta = 2t$$

where $t_p = \mu t_g$ is the optical spacing between the plates,

$$p(\theta, \lambda) = \lambda\phi/2\pi \cos\theta$$

and t is the effective thickness of the etalon spacer.

Constructive interference and maximum transmission occurs when m is integral. For an ideal etalon with perfectly flat and parallel plates the transmission of the etalon is given by the Airy function which, expressed in terms of m, has a period of unity:

$$A'(m) = \tau_A/[1 + F \sin^2 (\pi m)]$$

where $\tau_A = T^2/(1 - R)^2$ and $F = 4R/(1 - R)^2$ and T and R are the transmittance and reflectance of the plate surfaces.

If all-dielectric coatings are used their absorption and scattering are small i.e., $T \simeq 1 - R$ and $\tau_A \simeq 1$ and $A'(m) \simeq 1$ for m integral. If metal or metal-dielectric coatings are used their absorption may be significant, hence τ_A may be much less than unity. Although $p(\theta, \lambda)$ typically varies rapidly with λ in all-dielectric coatings, causing some difficulties in absolute wavelength calibration of the spectrometer, the advantage of greater etalon transmission is usually paramount for applications to airglow studies.

For usefully large values of R the Airy function is sharply peaked at integral values of m. The width of the peaks at half maximum of $A'(m)$ is

$$\delta_A(m) \simeq (1 - R)/\pi R^{1/2}$$

The quantity $N_A = \pi R^{1/2}/(1 - R)$, the ratio of peak spacing to peak width, is known as the reflective finesse of the coatings.

For a particular value of the effective plate spacing t the etalon transmission $A'(m)$ is a function of λ only, denoted $A(\lambda)$. Its period, known as the free spectral range, is

$$\Delta\lambda \simeq \lambda/m = \lambda^2/2t \quad \text{(for large m)};$$

the width of the peaks is

$$\delta_A(\lambda) = \Delta\lambda/N_A = \delta_A(m) \lambda/m$$

If now we select a small range of wavelengths, less than $\Delta\lambda$, including only the spectral feature of interest, with the aid of a narrow bandpass interference filter for example, and then vary t, the instrument becomes a scanning spectrometer. This may be done by varying t_g (separation scanning), μ (refractive index scanning) or θ (spatial scanning). Separation scanning is usually achieved by the use of piezoelectric supporting elements; refractive index scanning is usually achieved by variation of gas pressure in the etalon chamber (pressure scanning); spatial scanning might in principle be achieved with an annular field stop imaged by an optical system of variable power or by imaging the fringe (or fringes) onto a 2D imaging photon detector and measuring intensity versus θ, (radius in the image plane).

In an actual instrument the plates are not perfectly flat and parallel, the refractive index of the spacer medium may not be uniform, and nonuniformities

in the reflective coatings give rise to variations in $p(\theta,\lambda)$ across the aperture. For any small value of θ, the distribution of values of effective plate spacing about its mean t may be described by a "defect function" $D'(\bar{t} - t)$ such that $D'(\bar{t} - t)\ dt = dS/S$ is the fraction of the area of the aperture over which the effective plate spacing lies in the interval t to $t + dt$. Alternatively we may define a defect function in terms of wavelength so that

$$D(\bar{\lambda} - \lambda)d\lambda = dS/S$$

where $\bar{\lambda}$ defined by $m_r \bar{\lambda} = 2\bar{t}$ is the wavelength transmitted at mean effective spacing $\bar{t}$ and some reference value m_r of the order. Since

$$D(\bar{\lambda} - \lambda)d\lambda = D'(\bar{t} - t)\ dt$$

$$D(\bar{\lambda} - \lambda) = D'(\bar{t} - t)\ (m_r/2)$$

and $\delta_D(\lambda) = \delta_D(t)\ (2/m_r)$ is the width of the defect function. We define the defect finesse as the ratio

$$N_D = \Delta\lambda/\delta_D(\lambda) = (\lambda/2)/\delta_D(t)$$

$1/N_D$ is then the width of the defect function $D'(\bar{t} - t)$ in units of $\lambda/2$.

It should be noted that, in general, $D(\bar{t} - t)$ and $D(\bar{\lambda} - \lambda)$ are asymmetrical, i.e., $D'(\bar{t} - t) \neq D'(t - \bar{t})$ and $D(\bar{\lambda} - \lambda) \neq D(\lambda - \bar{\lambda})$.

The contribution to the etalon transmission of the fraction $D(\bar{\lambda} - \lambda)d\lambda$ of the aperture area is

$$dE = A(\lambda)\ D(\bar{\lambda} - \lambda)d\lambda$$

and the total transmission is

$$E(\bar{\lambda}) = \int A(\lambda)\ D(\bar{\lambda} - \lambda)d\lambda$$

or, in conventional notation,

$$E(\lambda) = A(\lambda) * D(\lambda)$$

i.e., the convolution of the Airy function and the plate defect function. $E(\lambda)$ is known as the etalon function.

The shape of the curve of $E(\lambda)$ versus λ depends on the shapes of both $A(\lambda)$ and $D(\bar{\lambda} - \lambda)$. Its width will be denoted $\delta_E(\lambda)$ and we define the etalon finesse

$$N_E = \Delta\lambda/\delta_E(\lambda)$$

To transmit power through the etalon a finite solid angle of view is necessary, i.e., a range of values of θ must be used, say θ_1 to θ_2 including the annular solid angle $\Delta\Omega = \Omega_2 - \Omega_1$. We then define the aperture function $F(\lambda_1 - \lambda)$, in the range $\lambda_2 \lesseqgtr \lambda \leq \lambda_1$ such that

$$-F(\lambda_1 - \lambda)\ d\lambda = d\Omega/\Omega$$

is the fraction of the total solid angle $\Delta\Omega$ within which the wavelength of maximum transmission lies in the interval λ to $\lambda + d\lambda$. Using the relations $\Omega = 2\pi\ (1 - \cos\ \theta)$ and $\lambda = \lambda_o \cos\ \theta$, where λ_o is transmitted at $\theta = 0$, we derive

$$F(\lambda_1 - \lambda) = 2\pi/\lambda_0 \Delta\Omega$$

which is constant in the range $\lambda_2 \leq \lambda \leq \lambda_1$ and zero outside that range. The width of the function $F(\lambda_1 - \lambda)$ is

$$\delta_F(\lambda) = \lambda_0 \Delta\Omega /2\pi$$

We define the aperture finesse by

$$N_F = \Delta\lambda/\delta_F(\lambda)$$

The transmission of the etalon with field of view $\theta_1 \leq \theta \leq \theta_2$ is then

$$I(\lambda_1) = \int E(\lambda) F(\lambda_1 - \lambda)d\lambda$$

or

$$I(\lambda) = E(\lambda) * F(\lambda)$$

This is the instrument function describing the transmission of the etalon for a fixed optical spacing and solid angle $\Delta\Omega$. Its shape depends on the shape of $E(\lambda)$ and the width of the aperture function. The width of the instrument function is denoted $\delta_I(\lambda)$ and we define the instrument finesse

$$N_I = \Delta\lambda/\delta_I(\lambda)$$

Commonly only one transmission fringe of the etalon is used but it is possible to use a multiple ring field stop transmitting several fringes. We may treat this as equivalent to the case of a single fringe field stop so long as $\Delta\Omega$ is the same for each fringe and they are spaced at integral intervals in m; also, in calculating the total light flux through the etalon we must replace $\Delta\Omega$ by $\Omega = n \Delta \Omega$ where n is the number of fringes viewed.

Finally we consider the etalon performance for light with a total radiance L and source function $B(\lambda)$ such that

$$B(\lambda)d\lambda = dL/L$$

is the fraction of the radiance within the wavelength interval λ to $\lambda + d\lambda$. In the present application we are concerned with line source profiles.

The output function of the etalon, i.e., the fraction of the total radiance transmitted, may now be written as

$$Y(\lambda_s) = \int I(\lambda_s - \lambda) \cdot B(\lambda)d\lambda$$

in which we have referred the instrument function $I(\lambda_s - \lambda)$ to the scanning parameter $\lambda_s = 2t_s/m_0$ where m_0 is the (fixed) order of interference at say, $\theta = 0$ and t_s is the (variable) effective plate spacing.

In the usual notation

$$Y(\lambda) = I(\lambda) * B(\lambda)$$

or

$$Y(\lambda) = A(\lambda) * D(\lambda) * F(\lambda) * B(\lambda)$$

It can be shown (Papoulis, 1968) that if

$$f_o(t) = f_1(t) * f_2(t) * \ldots * f_n(t)$$

and $\alpha_i = \int f_i(t)dt$ is the area under $f_i(t)$, then

$$\alpha_o = \alpha_1 \cdot \alpha_2 \ldots \alpha_n$$

and the variance of $f_o(t)$ is given by

$$\sigma_o{}^2 = \sigma_1{}^2 + \ldots + \sigma_n{}^2$$

In the present case the area under one period of the Airy function is

$$\alpha_A = \Delta\lambda \ T^2/(1 - R^2)$$

and, because of their definitions,

$$\alpha_D = \alpha_F = 1$$

Consequently,

$$\alpha_Y = \alpha_I = \alpha_E = \alpha_A$$

Taking the width at half maximum as a measure of σ,

$$\delta_E{}^2(\lambda) \simeq \delta_D{}^2(\lambda) + \delta_A{}^2(\lambda)$$
$$\delta_I{}^2(\lambda) \simeq \delta_E{}^2(\lambda) + \delta_F{}^2(\lambda)$$
$$\delta_Y{}^2(\lambda) \simeq \delta_I{}^2(\lambda) + \delta_B{}^2(\lambda).$$

The total light flux transmitted by the spectrometer is

$$\phi(\lambda) = \tau_c S \ L \ \Omega Y(\lambda)$$

where τ_c is the effective transmission coefficient of all components of the spectrometer other than the etalon and Ω is the total solid angle of the field of view of the etalon. Assuming the full free spectral range is recorded the total number of photons detected during the recording of a line profile is then

$$\Phi = \tau_c \epsilon \ SLT\Omega\alpha_A$$

where ϵ is the effective quantum efficiency of the detector and T is the duration of the recording. Using the expression previously derived for the width of the aperture function and noting that, for small absorption in the plate coatings and R in the range of present interests, $\alpha_A \simeq 1.5 \ \delta_A(\lambda)$, we obtain

$$\Phi \simeq 3\pi \ \tau_c \epsilon SLT \ \delta_F(\lambda)\delta_A(\lambda)/\lambda$$

CHOICE OF OPERATING PARAMETERS

In application to the airglow the source is very weak but slowly varying; the aurora is typically not so weak but varies rapidly. In either case it is desirable to optimize the choice of instrument parameters to minimize the uncertainty in the quantity being measured. A rough guide that is frequently used to choose $\delta_I(\lambda) \simeq \delta_B(\lambda)$ and maximize Φ. This implies $\delta_F(\lambda) \simeq \delta_A(\lambda)$ while $\delta_D(\lambda)$ is as small as feasible.

More rigorous and accurate criteria for optimization are discussed by GAGNE et al. (1974), HERNANDEZ (1978a, 1979) and JAHN et al. (1982).

Apart from these considerations the more obvious ones should not be over-looked. The effective transmission of the non-etalon components should be kept as high as feasible. Anti-reflection coatings should be used on all surfaces; if the instrument is to be used at a single wavelength two layer "V coatings" may be used, otherwise broadband multi-layer coatings are needed.

Obviously the higher the effective quantum efficiency of the detector the better but, especially for night airglow observations, it is important that the photocathode dark current be as small as feasible.

The higher the value of N_D the better but there is little to be gained by seeking values greater than $\sim$40 which is attainable with plates up to 150 mm diameter. It should be noted that non-uniformities in the reflective coatings introduce variations across the aperture in the phase change on reflection and can seriously reduce N_D. Also, there is not a wide range of designs available to achieve any desired value of N_A with broad-band stress free coatings.

MEASUREMENT OF ETALON CHARACTERISTICS

(a) Reflective Finesse, N_A

The manufacturer of the coatings should be able to provide data on N_A but variations may arise through aging. To measure N_A the following procedure may be used. Select:

1. Small plate spacing and hence small order m (which must be measured) and large $\Delta\lambda$.
2. Small aperture stop (near the plane of the etalon) to ensure that, within the aperture, N_D is very large.
3. Small field of view Ω and hence small $\delta_F(\lambda)$.
4. We now record the profile of several (narrow) lines at different wave-lengths in the region of interest.
5. Using the best estimates available for the various quantities involved we then use the relation

$$\delta_A^2(\lambda) \simeq \delta_Y^2(\lambda) - \delta_F^2(\lambda) - \delta_D^2(\lambda) - \delta_B^2(\lambda) \ .$$

to determine $\delta_A(\lambda)$ and hence $N_A(\lambda)$.

(b) Etalon and Defect Finesse, N_E and N_D

With the same procedure as above except that the full etalon aperture is used we then have

$$\delta_E^2(\lambda) \simeq \delta_Y^2(\lambda) - \delta_F^2(\lambda) - \delta_B^2(\lambda)$$

and

$$N_E \simeq \Delta\lambda / \delta_E(\lambda)$$

and

$$N_D^{-2} \simeq N_E^{-2} - N_A^{-2}$$

(c) Instrument Function $I(\lambda)$

The instrument function at any wavelength may be determined as follows using the normal operational values of λ, plate spacing and field of view.

1. Record the profile of a narrow line source at wavelength say λ_o e.g., a

single mode laser or the $\lambda546.1$ nm line of Hg-198. If the latter is used it will be necessary to deconvolve with the known line profile to determine the instrument profile $I(\lambda_o - \lambda)$. This will normally be obtained as a set of numbers corresponding to values of effective plate spacing t.

2. Deconvolve $I(\lambda_o - \lambda)$ with $A(\lambda)$ using the appropriate value of $R(\lambda_o)$ derived from $N_A^o(\lambda_o)$. This will yield $F(\lambda) * D(\lambda)$, the result being independent of wavelength when considered as a function of t.

3. Truncate or extend the tails of this data block to make the number of points correspond to one order at the new wavelength λ_1.

4. Convolve $F(\lambda) * D(\lambda)$ with $A(\lambda)$ using $R(\lambda_1)$ corresponding to the new wavelength λ_1. This will yield the instrument function $I(\lambda_1 - \lambda)$.

(d) Determination of Order and Wavelength Calibration

Using the normal operational values of plate spacing and field of view the profiles of several lines of known wavelength are recorded. If λ_r denotes one of these taken as a reference wavelength and λ_x denotes another, then for $\theta = 0$

$$m_r \lambda_r = 2 t_{pr} + 2 p_r$$

$$m_x \lambda_x = 2 t_{px} + 2 p_r$$

and

$$m_r \lambda_r - m_x \lambda_x = 2 (t_{pr} - t_{px}) + 2 (p_r - p_x)$$

where t_{pr} and t_{px} denote the optical spacings corresponding to the peaks of the line profiles and hence m_r and m_x are integral. The difference $(t_{pr} - t_{px})$ is measured precisely from the record of the two line profiles and, if separation scanning is used, taking account of the variation of μ with λ. Using the manufacturer's values for p_r and p_x (and making the reasonable assumption that the error in $2(p_r - p_x)$ is less than $\lambda_r - \lambda_x$) we can then determine m_r and m_x uniquely and hence obtain t_r, t_x and a revised value for p_x.

This procedure is repeated for other values of λ_x to determine $p(\lambda)$ and hence to revise the manufacturer's curve for $p(\lambda)$ versus λ. The same method may be used for any value of $\theta \neq 0$ to obtain $p(\theta,\lambda)$. The determination of any unknown wavelength is now, in principle, straightforward.

An alternative method of precise wavelength determination not requiring knowledge of the function $p(\theta,\lambda)$ is described by STANLEY and ANDREW (1964). This, however, necessitates measurements at two different values of plate spacing. As this usually takes a considerable time to achieve with the types of instrument used for the present applications, and as the precise wavelength varies with time, this method is not useful.

For determination of horizontal components of the wind it has been common practice to calibrate the spectrometer by pointing it vertically and assuming the vertical component of the wind to be zero. This is of no use if the prime objective is to measure the vertical component of the wind. For this purpose one may use the method described above to determine the mean value of $p(\lambda)$ from zenith observations extending over a long period of time and then calibrate individual observations by reference to a laboratory source for λ_r.

Finally, the most direct method of calibration involves comparison with a laboratory source of the emission line of interest. This may be used directly, or indirectly by determining $p(\lambda)$ so that another more convenient reference

wavelength λ_r may be used routinely. There is, of course, a difficulty when the emission of interest is from metastable atoms but laboratory sources of the OI $\lambda558$ and $\lambda630$ lines have been built (SHEPHERD, private communication and TOROP, private communication). Ultimately such laboratory sources will be essential for the determination of vertical winds because the zero velocity values of the wavelengths of the emission lines of interest have not yet been measured with sufficient accuracy to permit us to assign absolute values of vertical winds with adequate accuracy.

Probably the most generally convenient reference wavelength for the measurements described above is the $\lambda546.1$ mm emission from the single isotope Hg-198. This may be excited at adequate intensities for present purposes with a very simple UHF oscillator operating on less than 100 mW dc input power. The excitation is initiated with an auxiliary circuit and is maintained by the field between two plates which form part of the output tuned circuit of the oscillator (CREIGHTON – private communication). While a single mode laser is very satisfactory for determination of instrument profiles, a very high degree of stabilization is required if it is to be used as a wavelength reference for wind velocity determination.

DATA REDUCTION

To recover information on the source line profile shape and position, account must be taken of the line-broadening introduced by the instrument function. There are two main approaches to this task, different in detail but identical in principle.

The first approach is to deconvolve the instrument function from the recorded profile numerically to arrive at a representation of the source function, then to fit these data with some analytical function such as a Gaussian profile. The parameters of best fit give estimates of the line width, centre position and strength. Care needs to be taken in the fitting procedure that a suitable weighting scheme is used so that noise in the recorded profile, which is amplified by the deconvolution process, does not dominate the result. JACKA et al. (1980) used this method.

The second approach is to form a synthetic recorded profile by numerically convolving an analytical trial profile with the instrument function, and to adjust the parameters of the trail profile to give best fit between this convolution product and the actual recorded profile. This approach avoids the need for the weighting scheme mentioned above but may be more lengthy computationally unless the instrument function can be approximated analytically. HAYS and ROBLE (1971) have described a version of this technique.

In either case, and especially for recordings with many channels of data, there are advantages in performing the numerical convolutions and deconvolutions by multiplying or dividing the discrete Fourier transforms of the data, and doing the curve-fitting in the transform domain. Use of the discrete Fourier transform is also convenient in processing an instrument function recorded at one wavelength so that it is applicable at another, as described earlier.

CONTAMINATION OF RECORDS

Care should be taken to ensure that emission lines other than the one of interest are effectively blocked or excluded from the record by careful choice of the exact value of optical spacing of the etalon plates. For example, when observing the OI$\lambda558$ mm lines in the airglow the OH(7-1) band may cause trouble while the OH(8-2) band may be troublesome when observing the Na D lines. Reference should initially be made to an atlas of the spectrum such as that of BROADFOOT and KENDALL (1968) for the airglow or those presented by

VALLANCE-JONES (1974) for the aurora.

A more subtle problem inherent in ground based observations is the risk of contamination of the signal by light scattered from other regions of the sky; this may be especially serious in the auroral region (ABREU et al., 1983).

INSTRUMENT DESIGN

Some aspects of Fabry-Perot spectrometer design will be outlined by reference to the instrument at Mawson, Antarctica. Other instruments will be mentioned briefly.

(a) The Mawson (Antarctica) Spectrometer

This is a dual, separation scanning instrument in which the central fringe only is recorded. It was designed for application to the airglow and aurora during night and day and is useful over a broad range of wavelengths. It is a development from that described by JACKA et al. (1980) and COCKS et al. (1980).

The Mawson instrument is illustrated in Figure 5. Its operation is controlled electronically from a console external to the spectrometer housing.

The viewing periscope, driven in steps of 0.1 degree of zenith angle and 1 degree of azimuth angle by microprocessor controlled stepper motors, is mounted on the roof. A simple solar telescope on an equatorial mount, incorporated into the periscope, may be brought into use to permit subtraction of the scattered solar spectrum from daytime records.

The spectrometer, within a temperature controlled cabinet, is supported on three stainless steel bellows in tension, pressurised on the outside within an annular chamber; this support has a bounce period of ~1 sec.

The low resolution etalon may be swung horizontally out of the optical path for night time observing.

The collimator lenses are f/5.6 Taylor triplets designed by A. L. H. Aldersey and manufactured by Quentron Optics, Adelaide.

The fused silica plates in the low resolution etalon were made by Quentron Optics, Adelaide; those in the high resolution etalon were made by I. C. Optical Systems Ltd., London, and repolished by CSIRO, National Measurement Laboratory, Sydney. The coatings are the 13-layer all-dielectric type described by NETTERFIELD et al. (1980). These have acceptable (but somewhat high) reflectance over the wavelength range 450 nm to 720 nm. After ageing the reflective finesse was $N_A \simeq 27$ at $\lambda546$ nm and $N_A \simeq 33$ at $\lambda632$ nm. In the wavelength range 800 nm to 875 nm, which is used for parallelism control, the reflective finesse varies from about 60 to 20. The defect finesse of the plate in operation is $N_D \simeq 60$ for the low resolution etalon and $N_D \simeq 40$ for the high resolution etalon.

The etalon plates are mounted within "meehanite" rings which are supported by PZT 5H piezoelectric tubes polarized radially. Large variations in plate spacing are achieved by three steel support screws driven through worm drives by stepper motors.

Etalon plate parallelism across two orthogonal diameters is servo-controlled using the method of RAMSAY (1962) in which collimated, near infrared light passes down through the etalon space on one side and up through the opposite side to a PIN diode. White light superposition fringes are formed; the maximum signal indicates parallelism. The direction of the error is sensed by
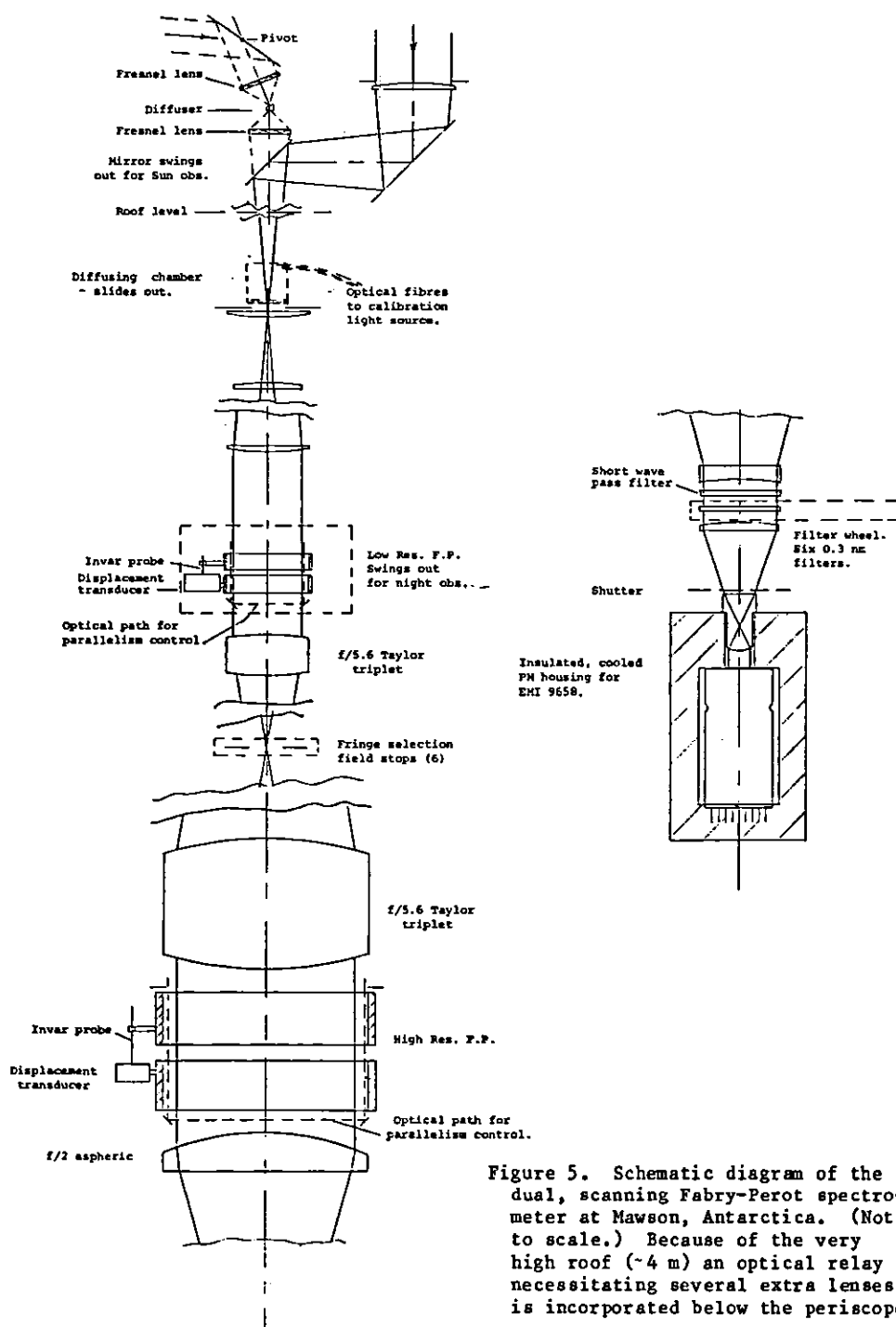
Figure 5. Schematic diagram of the dual, scanning Fabry-Perot spectrometer at Mawson, Antarctica. (Not to scale.) Because of the very high roof (~4 m) an optical relay necessitating several extra lenses is incorporated below the periscope.

applying a very small amplitude ~1 kHz three phase wobble to one plate and using phase sensitive detection on the output from the PIN diodes. Because of scattering some of the light passes down towards the photomultiplier necessitating the incorporation of a short wavelength pass filter in the main optical system.

Plate separation is servo-controlled to follow a stepped, saw-tooth wave form. It is measured with an air spaced capacitor displacement transducer; this is an improved version of that described by JACKA et al. (1980). The control voltage determines the channel "address" for the output from the photo-multiplier.

A pressure transducer senses the air pressure in the etalon chamber; its output is used to feed a correction to the plate separation to compensate for changes in refractive index of the air. The magnitude of this effect, in terms of geometrical spacing is:

$$\Delta t_g = (\mu_0 - 1) \; t_g \; \Delta p/p_0$$

Measurement of wind velocity with a precision of $\delta v = 5 \text{ ms}^{-1}$ requires

$$\delta v/c \simeq \delta\lambda/\lambda = (\mu_0 - 1) \; \delta p/p_0$$

or $\delta p \simeq 0.06$ mb.

The pressure transducer at present in use is a National Semiconductor (USA) LX1702A and is barely adequate for this precision. A Parascientific Inc. "Digiquartz" 230-AS-002 will soon be installed to see if the long term drift in wavelength calibration is improved.

An alternative approach would have been to enclose each etalon in a rigid air tight chamber.

To reduce drifts the temperature in the etalon chambers is controlled within about $\pm 5 \times 10^{-3}$°C. The probable principal effect of the residual temperature variations, which are not uniform throughout the chamber, is to cause distortion of the plate support structures. This effect could have been reduced (at considerable expense) by using invar rather than steel and meehanite for the structure.

Wavelength calibrations are made at least every 20 minutes; the uncertainty in wind velocity measurement due to instrument drift is then usually less than 10 ms^{-1}.

Plate spacing can be set at any value up to ~25 mm in the high resolution etalon and ~5 mm in the low resolution etalon. The piezoelectric supports permit scanning over a range of ~1200 nm. For routine observing the exact order is chosen so that the emission line of interest and a reference line (usually the Hg-198 λ546.1 nm line) are recorded near the middle of a scan which normally spans ~0.6 of a free spectral range. The scanning speed may be varied over a wide range but typically a period around 2 sec is used for the aurora and around 10 sec for the airglow.

For daytime observations the general principles outlined by COCKS et al. (1980) are followed. However, because of the much higher defect finesse and lower reflective finesse of the plates in the Mawson instrument and because both etalons are fully servo-controlled and their stability improved, the performance of the spectrometer is much better.

Calibration light sources can be introduced via a fibre bundle and

diffusing chamber above the upper field stop. A selection of 6 fringe selection stops is mounted on a "Maltese cross" mechanism. Up to six interference filters are carried in a wheel above the photomultiplier.

The photomultiplier is an EMI 9658A which has an S20 cathode on a multi-pyramidal form impressed into the inside surface of the front window. This increases the effective quantum efficiency of the photocathode. The photomultiplier is cooled to --30°C by a Products for Research TE-104TS-RF Peltier junction cooler. There are other photomultipliers capable of better performance for night observing, e.g., the RCA C31034A with a gallium arsenide chip cathode. However, because of the large signals involved, these are not suitable for daytime observing.

The data acquisition system comprises three microprocessors each performing separate functions but accessing a block of common memory. One processor displays accumulated data on a CRO or plotter, one operates the data recording facility, currently a digital cassette recorder, while the third controls data acquisition, including scan waveform generation, observing program, etc. The system has 256 channels, each with a maximum count capability of 2^{24}, which may be used as one block or two 128-channel blocks for accumulating data.

For night observing signal accumulation is by photomultiplier pulse counting; for daytime observing the photomultiplier current is amplified and a VCO is used to generate pulses for input to the pulse counting unit. Very good linearity of response is required for daytime observing because of the need to subtract the direct sun signal from the (normalized) sky signal (COCKS et al., 1980).

"Housekeeping" data including periscope zenith and azimuth angle, date, start and finish times, field stop number, filter number, number of scans and scan speed etc., are recorded along with the signal counts.

(b) Other Fabry-Perot Spectrometers

One of the most important separation scanning instruments in use is that originally described by HERNANDEZ and MILLS (1973). In this instrument plate parallelism is controlled by creating correcting signals proportional to the relative displacements of the fringe maxima of three Hg λ546 nm line sources spaced near the three piezoelectric supports. Stabilization of the spacing at a fixed point in the scan is provided by reference to the fringe maximum. Linearity of the scan is dependent on the wise use of the properties of the piezoelectric supports (HERNANDEZ, 1978b). More than one order is normally scanned.

A further important development proposed by HERNANDEZ et al. (1981) is a twin etalon scanning spectrometer (TESS). This involves the use of two similar etalons in series, one with fixed optical spacing t_1 and one variable t_2. The fixed etalon, illuminated by a (nearly) monochromatic source transmits fringes the maxima of which are defined by integral values of $m = 2t_1 \cos \theta/\lambda$, the total transmission being $I_1(\lambda) \cdot B(\lambda)$. If the second etalon is scanned by varying the optical spacing, all fringes will be transmitted but the output will be modulated, with maximum occuring when $t_1 - t_2 = 0$ and minimum when $|t_1 - t_2| = \lambda/4$. The total transmission is

$$Y(\lambda_s) = \int [I_1(\lambda) \cdot B(\lambda)] \cdot I_2(\lambda_s - \lambda) \, d\lambda$$

or

$$Y(\lambda) = [I_1(\lambda) \cdot B(\lambda)] * I_2(\lambda)$$

By using a TESS with 21 fringes HERNANDEZ et al. (1981) obtained a luminosity gain of 13.6 over that for a singe etalon with a single fringe. They estimate that gains of 50 are attainable.

An example of a pressure scanned instrument is that described by SIPLER et al. (1983), a development from that originally described by FEIBELMAN et al. (1972). This has a 94 mm aperture and views 6 fringes. Normally ~0.75 of a free spectral range is scanned in 25 sec. A frequency stabilized He-Ne laser scanned over 1.5 orders is used to establish the wavelength scale and a reference wavelength for wind measurements and to determine the instrument function.

It will be apparent from the comment above on pressure corrections in separation scanning etalons that if pressure scanning is used, with air or other gas of similar refractive index such as argon or nitrogen, the pressure must be monitored with a precision of ~0.06 mb in order to measure wind velocity with a precision of ~5 ms^{-1}. An important advantage of pressure scanning (and spatial scanning) is that fixed spacers, not requiring servo control, may be used. That also implies a lack of flexibility in applications requiring changes in free spectral range.

A spatial scanning instrument for use on Dynamics Explorer is described by HAYS et al. (1981), REES et al. (1982) and KILLEEN et al. (1982). This has a working aperture of 33 mm; fixed 12.6 mm spacers of Schott "Zerodur" are cemented in with cyanoacrylic adhesive. This instrument permits wind measurement with a precision of ~10 ms^{-1}.

A single fringe is imaged onto the photocathode of a 2D imaging photon detector, the anode of which is made up of 12 concentric rings, the output from which represents the fringe profile in 12 channels.

(c) Michelson Interferometers

A wide angle Michelson interferometer (WAMI) has been used by HILLIARD and SHEPHERD (1966) for measurement of Doppler widths of the λ558 nm line in the airglow and aurora. GADSDEN (1977) has used a similar instrument for measurement of Doppler shift of the λ558 nm airglow emission.

The WAMI permits quite short observing times but appears to be expensive to construct and difficult to use.

IMAGING FABRY-PEROT SPECTROMETERS

An important requirement for the future, especially for the study of details of temperature variations and wind in the upper atmosphere of the auroral regions, is a high luminosity, large field of view imaging system.

Imaging systems for this purpose have been discussed by SIVJEE et al. (1980) and by REES and GREENAWAY (1983). The latter authors discuss several different systems and propose an array of three stations to permit determination of vector winds over a large region of space.

The most promising instrument seems to be the TESS of HERNANDEZ et al. (1981) with an image of the sky in a plane near the etalons re-imaged onto a 2D photon detector. The spectral information is then obtained by recording the output, from each pixel or group of pixels, as a function of the scanning parameter of the etalon. This instrument could be coupled to a field compressor thus permitting rapid recording of the Doppler temperature and radial velocity with good spatial resolution over the whole sky.

ACKNOWLEDGEMENTS

REFERENCES

Abreu, V. J., G. A. Schmitt, P. B. Hays, J. W. Meriwether, C. A. Tepley and L. L. Cogger (1983), Planet. Space Sci., 31, 303.

Basedow, R. W. (1977), Ph.D. Thesis, University of Adelaide.

Bates, D. R. (1981), Planet. Space Sci., 29, 1061.

Born, M. and E. Wolf (1965), Principles of Optics, Pergamon Press.

Broadfoot, A. L. and K. R. Kendall (1968), J. Geophys. Res., 73, 426.

Chabbal, R. J. (1953), J. Rech. Centre. Nat. Rech. Sci. Lab. Bellevue (Paris) 24, 138. English translation, AERE Lib. Trans., 778, Harwell, U.K.

Chamberlain, J. W. (1961), Physics of the Aurora and Airglow, Academic Press.

Cocks, T. D., D. F. Creighton and F. Jacka (1980), J. Atmos. Terr. Phys., 42, 499.

Feibelman, W. A., R. D. Hake, D. P. Sipler and M. A. Biondi (1972), J. Geophys. Res., 77, 1869.

Gadsden, M. (1977), Paper read at IAGA/IAMP Joint Assembly, Seattle.

Gagne, J. M., J. P. Saint-Dizier and M. Picard (1974), Appl. Opt., 13, 581.

Gulledge, I. S., D. M. Packer, S. G. Tilford and J. T. Vanderslice (1968), J. Geophys. Res., 73, 5535.

Hays, P. B. and R. G. Roble (1971), Appl. Opt., 10, 193.

Hays, P. B., T. L. Killeen and B. C. Kennedy (1981), Space Sci. Inst., 5, 395.

Hernandez, G. (1966), Appl. Opt., 5, 1745.

Hernandez, G. (1970), Appl. Opt., 9, 1591.

Hernandez, G. (1971), Planet. Space Sci., 19, 467.

Hernandez, G. (1974), Appl. Opt., 13, 2654.

Hernandez, G. (1975), Geophys. Res. Letts., 2, 103.

Hernandez, G. (1978a), Appl. Opt., 17, 2967.

Hernandez, G. (1978b), Appl. Opt., 17, 3088.

Hernandez, G. (1979), Appl. Opt., 18, 3826.

Hernandez, G. and O. A. Mills (1973), Appl. Opt., 12, 126.

Hernandez, G., O. A. Mills and J. L. Smith (1981), Appl. Opt., 20, 3687.

Hilliard, R. L. and G. G. Shepherd (1966), J. Opt. Soc. Am., 56, 362.

Jacka, F., A. R. D. Bower, D. F. Creighton and P. A. Wilksch (1980), J. Phys. E. Sci. Instrum., 13, 562.

Jahn, H., G. Fellberg, B. Gladitz and M. Scheele (1982), J. Opt. Soc. Am., 72, 386.

Killeen, T. L. and P. B. Hays (1981), Appl. Opt., 20, 3314.

Killeen, T. L., P. B. Hays, B. C. Kennedy and D. Rees (1982), Appl. Opt., 21, 3903.

Kirchhoff, V. W. J. H. (1983), Geophys. Res. Letts., 10, 721.

Lawrence, G. M. and M. J. McEwan (1973), J. Geophys. Res., 78, 8314.

McEwan, M. J. and L. F. Phillips (1975), Chemistry of the Atmosphere, (Edward Arnold).

Meriwether, J. W., D. G. Torr, J. C. G. Walker and A. O. Nier (1978), J. Geophys. Res., 83, 3311.

Netterfield, R. P., R. C. Schaeffer and W. G. Sainty (1980), Appl. Opt., 19, 3010.

Papoulis, A. (1968), Systems and Transforms with Applications in Optics, McGraw-Hill.

Ramsay, J. V. (1962), Appl. Opt., 1, 411.

Rees, D., T. J. Fuller-Rowell, A. Lyons, T. L. Killeen and P. B. Hays (1982), Appl. Opt., 21, 3896.

Rees, D. and A. H. Greenaway (1983), Appl. Opt., 22, 1078.
Sipler, D. P. and M. A. Biondi (1978), Planet. Space Sci., 26, 65.
Sipler, D. P., M. A. Biondi and R. G. Roble (1983), Planet. Space Sci., 31, 53.
Sivjee, G. G., T. J. Hallinan and G. R. Swenson (1980), Appl. Opt., 19, 2206.
Smith, R. W., G. G. Sivjee, R. D. Stewart, F. G. McCormac and C. S. Deehr (1982), J. Geophys. Res., 87, 4455.
Stanley, R. W. and K. L. Andrew (1964), J. Opt. Soc. Am., 54, 625.
Thomas, R. J. and R. A. Young (1981), J. Geophys. Res., 86, 7389.
Vallance-Jones, A. (1974), Aurora, D. Reidel Publishing Co.
Wallace, L. and M. B. McElroy (1966), Planet. Space Sci., 14, 677.

SCANNING FABRY-PEROT SPECTROMETERS AVAILABLE FOR MEASUREMENT OF UPPER ATMOSPHERE TEMPERATURES AND WINDS

(based on data supplied by J. W. Meriwether, University of Michigan)

Station	Etalon Diameter	Scanning Method*	Detector Surface	Comment	Contact
Ann Arbor, MI USA	100 150 150	P P Sp	GaAs or S20	Designed for dayglow	P. B. Hays Univ. of Michigan
Arequipa, Peru	100	P	GaAs		M. A. Biondi Univ. of Pittsburgh, J. W. Meriwether Univ. of Michigan
Arecibo, Puerto Rico	150	P	GaAs		C. A. Tepley Arecibo Obs., M. A. Biondi Univ. of Pittsburgh, R. G. Burnside Univ. of Michigan
Calgary, Canada	150	Sp,P	S20		J. W. Meriwether Univ. of Michigan, L. L. Cogger Univ. of Calgary
College, Alaska	150 150 150	P P Sep	GaAs		C. J. Romick Univ. of Alaska, G. R. Swenson NASA, G. Hernandez NOAA

Location		Period	Detector	Notes	Investigator
Fritz Peak, Colorado, USA	150 150	Sep Sep	GaAs		G. Hernandez NOAA
Kiruna, Sweden	120 120	Sp	S20		D. Rees Univ. College, London
Skibotu, Norway	130	Sep	S20		R. W. Smith Ulster College
Madison, WI USA	150 150	P	GaAs		F. L. Roesler Univ. of Wisconsin
Spitsbergen, Svalbard	130	Sep	GaAs		R. W. Smith Ulster College
Sondrestrom, Greeland	100 100	Sp,P	S20		J. W. Meriwether Univ. of Michigan
Thule, Greenland	100	Sp,P	S20	To be installed Oct. 1984	T. L. Killeen Univ. of Michigan, J. W. Meriwether Univ. of Michigan
Albany, NY USA	150	P	S20		J. S. Kim Univ. of New York
Saskatoon, Canada	150	P	GaAs	Not yet operational	D. J. McEwen Univ. of Saskatoon
Laurel Ridge, PA, USA	150	P	GaAs		M. A. Biondi Univ. of Pittsburgh
Mt Zao, Japan	60	P	S20		H. Kamiyama and S. Okano Tohoku Univ.
Zvenigorod, USSR	50				Yu. L. Truttse USSR Acad. of Science
Yakitok, USSR					V. I. Ignatiev USSR Acad. of Science
Portable instrument	100	P	GaAs		M. A. Biondi Univ. of Pittsburgh
Beveridge, Vic Australia	150	P	S20		P. L. Dyson LaTrobe Univ., Vic, Australia

Mt. Torrens, S.A., Australia	50 150	Sep	S20	Dayglow + nightglow observations	F. Jacka Univ. of Adelaide, S.A. Australia
Mawson, Antarctica	50 150	Sep	S20	Dayglow, nightglow + aurora observations	F. Jacka Univ. of Adelaide, S.A. Australia
Halley, Antarctica	details not known				R. W. Smith Ulster College

* P   = pressure scanning
  Sep = separation scanning
  Sp = spatial scanning

3. OPTICAL STUDIES OF NITROGEN OXIDES IN THE STRATOSPHERE

J. F. Noxon

Fritz Peak Observatory
NOAA Aeronomy Laboratory
Boulder, CO

This discussion will be rather brief and mainly an overview with references, since the methods and results are all accessible in recent detailed publications; these also contain references to earlier work.

Several observational approaches have been utilized to study the oxides of nitrogen in the stratosphere; the two species accessible in the visible are NO_2 and NO_3 which absorb, respectively, in the blue (~400-450 nm) and red (620-670 nm). In the infrared NO, NO_2, and HNO_3 can both be studied easily only if measurements are made from above the tropopause where the water density becomes low. NO has been investigated in the mesosphere with rockets and satellites using ultraviolet absorption, but this will not be discussed further here.

Measurements have been carried out both by ground-based techniques as well as aircraft and balloons. Generally the sun or moon has served as the light source although there have been a few balloon measurements in which a telescope was employed in order to image a bright star or planet upon the slit of the spectrometer. A considerable degree of spectral selectivity is needed in order to positively identify a particular absorber; some of the earlier work on NO_2 has proved to be erroneous simply because rather broad filters were employed.

Ground-based studies of stratospheric NO_2 are discussed in the papers by NOXON et al. (1983) and colleagues, SYED and HARRISON (1980), and McKENZIE and JOHNSTON (1983). To a large extent this work utilizes measurements of NO_2 absorption in the zenith sky during twilight since a large geometric enhancement in the effective air mass may be obtained. It is also possible to make good measurements of the total column abundance of NO_2 using either the sun or moon when they are near the horizon (and the troposphere is relatively clean). Stratospheric NO_3 at night can be measured (NOXON et al., 1978), but it is not clear that N_2O_5 has been.

Aircraft measurements have been reported by BREWER et al. (1973), NOXON et al. (1978, 1983), COFFEY et al. (1981) and GIRARD et al. (1983) using both visible and infrared absorptions. The infrared study by COFFEY et al. (1981) and GIRARD et al. (1983) measured NO, NO_2 and HNO_3 as a function of latitude.

Balloon measurements of NO, NO_2, HNO_3 and NO_3 are discussed by MURCRAY et al. (1978), RIGEAUD et al. (1983), and EVANS et al. (1982). These are particularly valuable for their information on the altitude profiles. Recent work by MOUNT et al. (1983) reports satellite measurements of NO_2 in the stratosphere.

REFERENCES

Brewer, A. W., C. F. McElroy and J. B. Kerr (1973), Nitrogen dioxide concentration in the atmosphere, Nature, 246, 129.

Coffey, M. T., W. G. Mankin and A. Goldman (1981), Simultaneous spectroscopic determination of the latitudinal, seasonal and diurnal variability of stratospheric N_2O, NO, NO_2 and HNO_3, J. Geophys. Res., 86, 7331.

42

Evans, W. et al. (1982), Simulation of the October 23, 1980 stratoprobe flight, Geophys. Res. Lett., 9, 223.

Girard, A. et al. (1983), Latitudinal distribution of the stratospheric species deduced from simultaneous spectroscopic measurements, J. Geophys. Res., 88, 5377.

McKenzie, R. L. and P. V. Johnston (1983), Stratospheric ozone observation simultaneous with NO_2 at 45°S, Geophys. Res. Lett., 10, 337.

Mount, G. et al. (1983), Measurements of NO_2 in the Earth's stratosphere using a limb scanning visible light spectrometer, Geophys. Res. Lett., 10, 265.

Murcray, D. et al. (1978), WMO Symposium, Toronto.

Noxon, J. F., R. B. Norton and W. R. Henderson (1978), Observations of atmospheric NO_3, Geophys. Res. Lett., 5, 675.

Noxon, J. F., R. B. Norton and W. Henderson (1983), Stratospheric NO_2: 3, The effects of large scale horizontal transport, J. Geophys. Res., 88, 5240.

Rigeaud, P., J. P. Naudet and D. Huguenin (1983), Simultaneous measurements of vertical distributions of stratospheric NO_3 and O_3 at different periods of the night, J. Geophys. Res., 88, 1463.

Syed, M. Q. and A. W. Harrison (1980), Ground based observations of stratospheric nitrogen dioxide, Canadian J. Phys., 58, 788.

See also many useful references in "The Stratosphere, 1981": WMO Report No. 11 available from WMO or NASA, Code 963, Greenbelt, MD

4. MICROWAVE RADIOMETRIC STUDIES OF COMPOSITION AND STRUCTURE

J. J. Olivero

Department of Meteorology
The Pennsylvania State University
University Park, PA

INTRODUCTION

The possibility of composition measurements in the middle atmosphere by ground-based microwave radiometry was pointed out in pioneering papers by BARRETT and CHUNG (1962) and by CROOM (1965). Unfortunately a decade of unsuccessful attempts were to precede the eventual confirmation of this measurement technique.

Microwave radiometry is the precise measurement of small amounts of electromagnetic radiation at quite short wavelengths (submillimeter to centimeter). The spectral intensity of the radiation field sensed will usually be structured by a single, isolated molecular emission/absorption line, resolved in great detail. Each molecular line provides atmospheric composition information. Atmospheric structure measurements, on the other hand, refer to the distributed fields of pressure and density which give rise to circulations. These are derivable from temperature measurements which can be accomplished from ground-based microwave radiometry. The most obvious example is the thermal emission of an atmospheric constituent that has a known distribution in the homosphere – N_2, O_2, Ar, CO_2 – of which O_2 has a number of useful microwave lines.

The growth of this field in the past three decades has been carefully nurtured by the radio astronomy community. Scientists working in the middle atmosphere have begun adapting these techniques to their observations largely within the past decade. Several groups are presently active worldwide.

(a) Advantages for Remote Sensing

Any candidate measurement technique can be evaluated in terms of its advantages and disadvantages. This section is designated to stress the advantages of microwave radiometry, nevertheless a few less advantageous aspects will be included at the end.

Ground-based observations of middle atmospheric composition are usually limited by tropospheric visibility. e.g., ozone measurements based on transmission of solar UV to the surface. The microwave has a great advantage over shorter wavelength techniques in that measurements can be made in spite of cloud cover, tropospheric and stratospheric aerosols, and even while snowing, and must be curtailed only during active rain episodes. The microwave is also a much less cluttered part of the spectrum than say the infrared. Most features are single, isolated rotational lines which overlap other features only in their far wings.

The structure of each microwave line is also rather uncomplicated because pure pressure (collisional) broadening prevails from the earth's surface well into the mesosphere. Nor are microwave lines subject to non-LTE (local thermodynamic equilibrium) effects at these altitudes, effects which plague infrared measurements in the upper stratosphere and throughout the mesosphere. (Microwave transitions undergo non-LTE effects in the thermosphere.) Microwave radiometer – spectrometers can easily resolve each line element in exquisite detail. Frequency resolution is practically unbounded; any line can be decomposed into 1000 or more parts.

Only modest vertical measurement resolution (8-10 km) is attainable from the ground, however this is comparable to ground-based infrared and ultraviolet techniques (except LIDAR), and to nadir viewing satellite sounders.

Each receiver system is limited in frequency coverage (without modification) thus line to line tuning is very limited and most systems are developed to look at a single line only. For most lines, useful signal to noise ratios are obtained with integration periods of order minutes; a few weak lines (e.g., 22 GHz H_2O line in the middle atmosphere) will require periods of integration of hours to tens of hours.

(b) Impact on Middle Atmosphere Studies

The network of ground-based Dobson ozone spectrophotometers taking careful observations over several decades has produced the basic framework upon which our understanding of natural variations in ozone abundance is based. Ground-based microwave radiometry has the capacity to do the same job, at least as well, with a really "all weather" capability. Such measurements should give improved vertical resolution, will extend well into the mesosphere, and will not depend upon external calibration. (All measurement systems must be accurately calibrated, however in the microwave various thermal loads are easily fashioned to fill this need.)

Easily automated, hands-off operation is rather straightforward which gives one the capability of round-the-clock observation with temporal resolution as fine as minutes. One can finally investigate the diurnal ozone cycle as well as effects due to gravity waves, planetary waves, equatorial waves, reversals of the zonal winds, and various other solar and geophysical events in time scales from minutes to years. Such data sets perfectly complement the quasi-global to global data sets available from aircraft and satellite remote sensing.

Besides observations of ozone itself, other trace constituents which are intimately interrelated to ozone photochemistry, especially H_2O, ClO, and CO, can be routinely and simultaneously observed in the same measurement volume over the same temporal and spatial scales. With the addition of a temperature profile measurement, from O_2 thermal emission in the microwave, one can plan elegant experiments to test photochemical and transport effects on trace gas composition (especially ozone).

RADIATIVE TRANSFER

Microwave radiation power, P, confined to an (equivalent rectangular) frequency bandwidth, B, is directly proportional to an absolute temperture, T, through Nyquist's Theorem

$$P = kTB$$

where k is Boltzmann's constant. At these frequencies blackbody radiation is very nearly linear in absolute temperature (the Rayleigh-Jeans Law applies) so that the Planck function, $B_\nu(T)$, may be approximated as

$$B_\nu(T) = 2kT\nu^2/c^2$$

where ν is the frequency and c is the speed of light. Thus physical objects emit microwave radiation at rates proportional to their (brightness) temperatures and any flux of radiation received is equivalent to observing a corresponding uniform field of view at brightness temperature, T_B. Sources vary from the 3 K of "empty" space to tens of thousands of degrees K for the sun. The general equation of radiative transfer for a ground-based observation at microwave frequencies is

$$T_B(\nu,o) = T_B(\nu,\infty)\ \exp[-\tau_\nu(o,\infty)] + \int_0^{\tau_\nu(o,\ell max)} T(\ell)\ \exp[-\tau_\nu(o,\ell)]d\tau_\nu(o,\ell)$$

where $T_B(\nu,\infty)$ is a source external to the atmosphere (e.g., sun or moon), $\tau_\nu(0,\infty)$ is the optical depth along the actual atmospheric path from levels 0 to ∞, the integral gives the contribution from atmospheric thermal emission which is everywhere proportional to the local absolute temperature, $T(\ell)$, and ℓ is the slant path length from the ground to a specific level, $\ell(z)$, which has an upper boundary (for computational purposes) of $\ell_{max}(z)$, and z is the altitude. For elevation angles $\alpha \cong 10$ degrees or more, $\Delta\ell \cong \sec(90-\alpha)\Delta z$.

The optical depth at frequency ν from the surface to level $\ell(z)$ is given as

$$\tau_\nu(o,\ell) = \int_0^\ell \sum_i k_\nu^i[p(\ell),\ T(\ell)]d\ell$$

or

$$\tau_\nu(o,\ell) = \int_0^\ell \sum_i \sigma_\nu^i [p(\ell),\ T(\ell)]n^i(\ell)d\ell$$

where the superscripts, i, denote the various microwave absorber/emitter species at frequency ν, $k_\nu[p,T]$ is the absorption coefficient (cm^{-1}), $\sigma_\nu[p,T]$ is the absorption cross section (cm^2), and n^i is the concentration of the ith constituent (cm^{-3}).

Fortunately only one of the two terms in the radiative transfer equation above is generally of any consequence for a particular experiment. Referring to Figure 1, one may choose to view the atmosphere itself (left side — labelled "EMISSION") or to view an external source, such as the sun (right side — labelled "ABSORPTION"). With a brightness temperature of order 10,000 K the solar disk easily overwhelms an atmospheric brightness temperature of order 100 K (300 K maximum). In Figure 1 the portions of the spectra at the bottom with the horizontal shading are meant to indicate the frequency ranges corresponding to interaction with the middle atmosphere, specifically to the shaded layer within the lower mesosphere shown in the upper parts of the figure. Useful discussions of microwave radiative transfer can be found in WATERS (1976) and in ULABY et al. (1981).

The microwave signal received at the surface contains information from all points along the atmospheric path $0 \leq \ell(z) \leq \infty$. In (thermal) emission, the signal amplitude is portional to the product of $T(\ell)$ and $n^i(\ell)$ at each level while the line width generated at that level is proportional to the pressure (and weakly on $T(\ell)$ as well). If $n^i(\ell)$ is well known (e.g., O_2) then $T(\ell)$ information may be extracted. If $n^i(\ell)$ is unknown, $T(\ell)$ must be supplied. In practice, climatological data on $T(\ell)$ is usually sufficient because the uncertainty in $n^i(\ell)$ due to an error in $T(\ell)$ is modest. (A ten percent error in temperature would correspond to 25 K out of 250 K, for example; this level of accuracy should be routine.) In an absorption measurement the dependence on $n^i(\ell)$ is more straightforward and the role of $T(\ell)$ is reduced considerably.

The relationship between the observed frequency components of the brightness temperature $T_{B,\nu}$ and the unknown constituent profile $n^i(\ell)$ can be cast in the generalized equation of remote sensing (see TWOMEY, 1977)

$$T_{B,\nu} = \int_0^{\ell max} g(\nu,\ell)\ n^i(\ell)d\ell$$
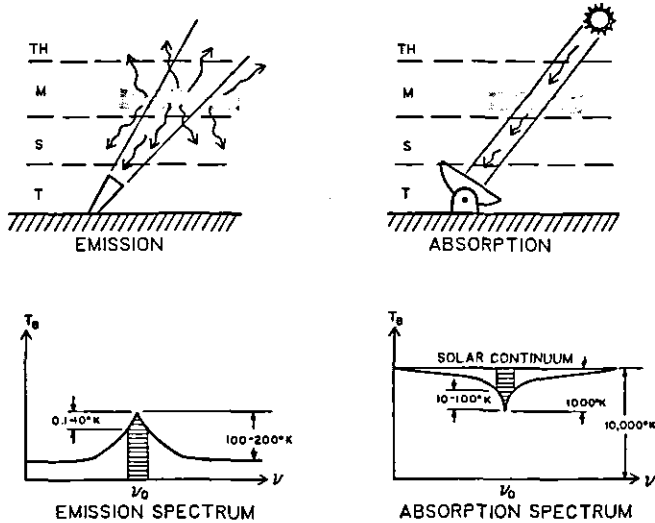
or

$$T = G\ N$$

Figure 1. Ground-based microwave measurement configurations;
emission and absorption. Top: source - observer geometry.
*Bottom: spectral line observed, horizontal stripes denote
principal contribution of mesospheric layer.*

where g(ν, ℓ) are called the kernels, contribution functions, or weighting
functions (they transpose height information into frequency information), T is
the observed vector of brightness temperatures at m frequencies. N is the
unknown constituent profile vector made up of r level values of $n^1(\ell)$, and G
is the m x r matrix for m simultaneous equations in r unknowns. The solution,
when mathematically feasible, would be simply

$$N = G^{-1} T$$

where G^{-1} is the inverse matrix of the weighting functions. BEVILACQUA et al.
(1983) discuss the inversion of 22 GHz water vapor data showing both the
practical limits on the information retrievable, and the propagation of errors.
(An even more complete discussion of inversions of microwave observations can be
found in BEVILACQUA, 1982.)

MICROWAVE RADIMETER SYSTEMS

The simplest microwave radiometer system consists of the following: an
antenna to collect and concentrate minute quantities of electromagnetic
radiation; a receiver to produce an output potential, V_{out}, proportional to
the input radiation (signal); and a spectrometer to spectrally analyze that
signal and further process, calibrate, and archive the data. In Figure 2 we see
a schematic drawing of a typical microwave radiometer system (taken from THACKER
et al., 1981); it may be useful to refer to this drawing throughout the
remainder of this discussion.

(a) Antenna

Antennas are devices which view a small portion of the environment, collect
most of the (microwave) electromagnetic radiation within the solid angle viewed,
and output that radiation power as an antenna temperature, T_A. The difference
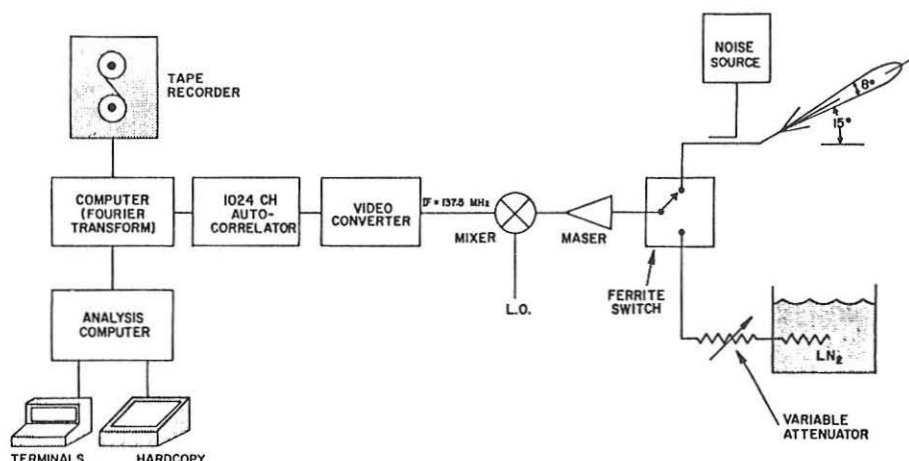
Figure 2. Schematic drawing of a microwave radiometer system
featuring a simple horn antenna (upper right), a noise
source periodically injected into the signal stream, calibration
and system balancing using liquid nitrogen (LN$_2$), a cryogenic
MASER amplifier, and conventional SSB heterodyne receiver -
back end (from THACKER et al., 1981).

between T_A and T_B, the effective brightness temperature viewed, corresponds
to the finite gain and losses in the antenna system. Careful calibration
procedures will establish this relationship. Referring to Figure 1, in emission
measurements the antenna requirements are straightforward in that one wishes to
gather radiation from a moderate beamwidth (say 5 to 15 arcdegrees in full
width). Such antennas have two major roles which are to define an elevation
angle and to suppress sidelobes - radiation sensed from far outside the main
beam. Rectangular and circular horns, either moveable in elevation or used with
reflectors, are commonly used in emission measurements.

Absorption measurements greatly increase the complexities of antenna
systems. Both the sun and the moon subtend angles of order 1/2 arcdegree. To
achieve beamwidths of this size or smaller one must use antennas such as para-
bolic or spherical sections with diameters of order meters. Moreover these
large antennas must accurately follow the source at rates of about 1/4 arc-
degrees per minute across the sky.

(b) Receiver

Radiometric receivers process a portion of the thermal noise-like power
(T_A) input to them in order to detect and suitably amplify the signal
contained. Virtually all such systems employ heterodyning a simple process for
producing intermediate frequency (IF) signals from the incoming microwave radio-
frequency (RF). A key component of the heterodyne receiver is a well stabilized
local oscillator (LO) signal source usually phase locked to a primary or
secondary frequency standard. A mixer is used to combine the RF and LO sources
into sum and difference beat frequencies. A conventional heterodyne system will
use the difference frequency (the lower one) as the IF. Two different RF
frequencies will mix down to the same IF frequency (LO+IF and LO-IF). If
preselection of only one RF frequency is used or if a special "image rejection"
mixer is used, only one of the two possible RF frequencies reaches the IF, and a
single side band (SSB) system results. The more simple double side band (DSB)

system produces images of both RF frequencies in the IF, where they become indistinguishable. The spectral analysis of such double side band systems must make some assumptions about the (ideally flat) behavior of the spectrum of the undesired "image response". Second and even third stages of heterodyning are sometimes used; the key being to obtain an IF sufficiently low that final amplification and processing are easy and efficient.

Power gains required often exceed 100 dB (10^{10}), but equally important is the requirement to minimize the noise added to the signal during processing. The receiver (noise) temperature is a measure of the idealness of the receiver. The output of an "ideal" receiver would be the amplified signal alone. Most observed signals can be characterized as an effective noise temperature, such as T_A. Unfortunately the finite physical temperature of the receiver components such as the mixer and amplifier produces noise, which is added to the desired signal. System non-linearities, reflections, standing waves, and other phenomena, which depend on the internal characteristics of the device, contribute to increasing the noise in the output. The receiver (noise) temperature is an effective input term, which when multiplied by the system gain results in the additional output noise observed. Carefully designed systems, especially those using cool parametric amplifiers and MASERs, can achieve receiver (noise) temperatures of 100 K or less.

A figure of merit for a particular measurement is the system (noise) temperature T_{SYS}, the sum of the receiver (noise) temperature and the antenna temperature. The system temperature is crucial in establishing the minimum detectible (signal) temperature, ΔT

$$\Delta T = T_{SYS}/\sqrt{B\tau}$$

where B is the bandwidth and τ is the integration period in seconds. The smaller the value of ΔT, the larger the actual signal to noise ratio for the measurement. If T_{SYS} is large then B and especially τ must be greatly increased to compensate. In atmospheric (thermal) emission measurements the receiver temperature is always important and will always dominate T_{SYS} for uncooled receivers. However, for solar absorption measurements, T_A (proportional to T_{SUN}) clearly dominates T_{SYS} and low receiver (noise) temperatures are of little benefit.

(c) Spectrometer

Microwave spectral line receivers, especially those used to probe the terrestrial middle atmosphere, are careful to preserve the molecular emission or absorption line shape to the maximum extent possible. Analysis of this line shape information is accomplished by the spectrometer system. The tried and true spectrum analyzer from radio astronomy studies has been the multichannel filter bank (PENFIELD, 1976). The rise of digital information processing over the past decade has increased the use of autocorrelator spectrometers (COOPER, 1976). Finally, military signal processing requirements are bringing to the forefront the LASER/solid state technology of the acoustic wave (SAW) or bulk acoustic wave (BAW) mode. See, for example, BERG et al. (1982a,b) and FUNG et al. (1982). In all cases, the goal is to provide a measure of the energy in many narrow spectral windows clustered around line center. The effective frequency resolution needed depends on the spectral line under study. Since the bandwidth, B, in the expression for the minimal detectable (signal) temperature is the bandwidth of each these narrow spectral windows, a compromise between frequeny resolution and observation time required must be made.

Also of great importance are the dedicated microprocessors and microcomputers used to run the experiment and to process the data. Radiometer systems require both periodic absolute calibration and continuous inter-

comparisons of source to reference (thermal) loads (in order to minimize the
effects of short-term gain variations). This can be done by periodically
changing the frequency of the local oscillator so that an adjacent frequency
band of the source is observed (frequency switching). Another method is to
periodically switch the receiver or antenna from the unknown source to a known
thermal reference (Dicke switching).

The output of the spectrometer channels must be accumulated and archived
along with a number of "housekeeping" parameters. Finally the data stream must
be processed to add calibrations and remove baselines (zero level offsets).
Actual inversion of the data into profiles is sometimes automated as well;
however, most of the observations listed in Table 1 used separate "hands-on"
inversion algorithms.

A BRIEF HISTORY OF MIDDLE ATMOSPHERE MEASUREMENTS

In Table 1 we list the ground-based microwave measurements of composition
and temperature which have already been made, along with some pertinent
information. Almost all such measurements have been made at midlatitudes, 30 to
45 degrees north, on the North American continent. In a sense this work has
been largely feasibility or proof of concept demonstration with location,
seasonal selection, etc., being of secondary concern.

(a) Ozone

This molecule is at the center of our concerns for anthropogenic altera-
tion of the middle atmosphere. As Table 1 shows, microwave measurements of
ozone have spanned the past decade. These measurements, as a whole, tend to
support photochemical models of ozone mixing ratio profiles. In addition,
they offered early, strong support for the diurnal variations of ozone,
especially in the mesosphere.

(b) Water Vapor

Ground-based microwave radiometric measurements of middle atmosphere water
vapor have been especially sought-after due to the extreme difficulty in
carrying out most other types of measurements. (Satellite infrared measurements
in the upper stratosphere are only now becoming generally available.) The HO_x
chemical family, derived from H_2 photolysis, is most intimately involved in
ozone photochemistry above about 30-40 km. Microwave measurements not only
support the general atmospheric profile deduced from photochemical models, but
also suggest secular variability on various time scales. This may be the
response of water vapor as a quasipassive tracer to atmospheric dynamics,
especially waves, on many temporal and spatial scales.

(c) Carbon Monoxide

Virtually the only measurements of mesospheric CO available today have come
from the ground-based microwave observations cited in Table 1.

(d) Chlorine Monoxide

The microwave measurements of ClO given in Table 1 have agreed in general
with in situ measurements (see ANDERSON et al., 1977, for example) as well as
photochemical models of the Cl_x - "odd oxygen" families (see FROIDEVAUX,
1983). There has been a single serious challenge to this point of view (see
MUMMA et al., 1983) and the issue remains unresolved at present.

(e) Perhydroxl

Stratospheric HO_2 has been observed in a single microwave study by

Table 1.

Molecule	Frequency (GHz)	Mode	Antenna Type	Antenna Elevation Angle	Receiver Bandwidth (MHz)	Spectrometer Resolution	Sensitivity	Integration Period (s)	Altitudes Covered (km)	Latitude	Site Elevation (m)	Comments	Reference
O_3 "	111 "	emission solar absorption	11m dish "	41° solar tracking	60 "	1.2 MHz "	$T_{SYS} \approx 3600°K$ "	10m (?)	0-80(?) "	32° "	1920 "	in radome "	Shimabukuro and Wilson (1973)
O_3	111	solar absorption	11m dish	solar tracking	100	1.2 MHz	$T_{SYS} \approx 3600°K$	10m	5-50	32°	1920	in radome	Shimabukuro et al. (1975)
O_3	111	emission	5m dish	30°	80	250 kHz	--	1h,1.5h	35-80	31°	2080	--	Penfield et al. (1976)
O_3	111	emission	2.5m dish	30°	25.6	100 kHz	$T_R = 350 \pm 50°$	9h,14.5h	~40-90	45°	73	20°K liquid helium cooled receiver	de La Noe et al (1983)
O_3 "	102 "	emission solar absorption	1.5m dish "	15° to 90° "	50 "	10 MHz	--	?	?	?	?	--	Caton et al. (1968)
O_3	102	emission	4.6m dish	~5°	64	250 kHz	$T_{SYS} = 500°K$	~1h	10-80	34°	38	--	Shimabukuro et al. (1977)
O_3	102	emission	4.6m dish	~10°	8	250 kHz	--	15m,2h,10d	40-70	34°	38	--	Wilson and Schwartz (1981)
O_3	142	emission	--	--	---	---	--	--	25-60	--	--	--	Lobsiger et al. (1983)
H_2O "	22 "	emission solar absorption	37m dish "	10° to 30° 1° to 13°	2 and 6.67	50 kHz "	$T_{SYS} = 150 \pm 30$ $T_R \approx 1900°K$	92.h ~50m	50-80 "	42.5° "	150 "	cryogenic MASER in radome	Radford et al. (1977)
H_2O	22	emission	8° horn	~15°	6.67	16 kHz	$T_{SYS} \approx 130°K$ $T_R \approx 80°K$	12h,3 to 7d	50-85	42.5°	150	cryogenic MASER	Thacker et al. (1981)
H_2O	22	emission	8° horn	~15°	6.67	16 kHz	$T_{SYS} \approx 130°K$ $T_R \approx 80°K$	100m	50-85	42.5°	150	cryogenic MASER	Gibbins et al. (1982)

H_2O	22	emission	8° horn	∿15°	6.67	16 kHz	$T_{SYS} \approx 130°K$ $T_R = 80°K$	12h, 3 to 7d	50-85	42.5°	150	cryogenic MASER	Bevilacqua et al. (1983)
H_2O	22	solar absorption	2.4m dish	solar tracking	2.5	50 kHz	$T_{SYS} \approx 12,000K$	4h to 8h	55-80	41°	360	--	Croskey et al. (1983)
H_2O	22	emission	8° horn	∿12°	40	--	--	110m	35-85	42.5°	150	cryogenic MASER	Schwartz et al. (1983)
CO	115	solar absorption	4.6m dish	solar tracking	8	250 kHz	--	5h	40-80	34°	38	--	Waters et al. (1976)
CO "	115 230	solar absorption "	6m dish "	solar tracking "	1.9 "	15 kHz "	--	4h "	50-80 "	41° "	1050 "	--	Goldsmith et al. (1979)
CO "	115 230	emission "	7m dish "	45° "	6.4 12.8	100 kHz 50 kHz	$T_R = 425°K$ $T_R = 3200°K$	155m 80m	50-80 "	41° "	200m "	--	Kunzi and Carlson (1982)
CO "	115 230	lunar absorption "	10.4m dish "	lunar tracking (day and night)	100 10	200 kHz 195 kHz	-- --	∿5.5h ∿5h	50-105 "	37° "	1216 "	acousto-optical spectrometer	Clancy et al. (1982)
CO "	230 "	emission lunar absorption	11m dish "	∿30° lunar tracking	2.56 and 64	100 kHz and 250 kHz	-- --	1h 1.5h	50-105 column above 65	32° "	1920 "	--	Clancy et al. (1982)
ClO	204	emission	6.3° horn	"low"	256	1 MHz	--	∿6h	25-50	42°	306	teflon lens	Parish et al (1981)
ClO	279	emission	6.3° horn	--	256	1 MHz	$T_R = 1100°K$	2h to 8d	25-60	20°	4200	teflon lens cryogenic mixer	Solomon et al. (1983)
HO_2	266	emission	--	--	256	1 MHz	$T_R = 1100°K$	∿9h	25-70	20°	4200	3 lines cryogenic mixer	deZafra et al. (1983)
O_2	52-54	emission	10° horn	90° (?)	40	160 kHz	$T_R = 1800°K$	16m to 4h	30-70	42.5°	150m	temperature	Waters (1973)

52

DEZAFRA et al. (1984). This is one of the HO_x radicals which link the ozone
photochemistry with water vapor (and its photolysis rate). Continued microwave
measurements will help evaluate rival photochemical models which bear directly
upon our ozone depletion concerns.

(f) Temperature

     As pointed out in the introduction, the emission measurement contains
information about the product of species abundance and local (kinetic)
temperature. Molecular oxygen, whose abundance is well known in the lower and
middle atmospheres, is an excellent candidate for temperature measurements.
Table 1 reveals the studies to date. Oxygen emission radiative transfer is
however greatly complicated both by the extensive overlapping and interference
among lines in the 50-60 GHz complex, and by the Zeeman splitting effect which
becomes important in the mesosphere.

THE NEAR FUTURE

     This field has grown as technology (and actual equipment) transfer from
radio astronomy to the atmospheric sciences has proceeded. There are two
obvious needs which nevertheless still must be addressed within the time frame
of the remainder of the initial MAP program period and the years immediately to
follow. The first is the acquisition of microwave radiometers dedicated to
atmospheric science. Virtually all measurements to date have used radio
astronomy observatory facilities as rather low priority guest users. Thus past
measurement programs have lacked continuity and in many cases lacked choice of
observing date at all.

     A second need is the extension of the observations to other latitudes and
to other continents. This will be done through establishing new programs abroad
and through the realization of the need for truly portable radiometer systems.
Thus continuous (microwave) composition and structure measurements might be
configured at any given field experiment location to support a major rocket
campaign or MST (mesosphere-stratosphere-troposphere) radar site, for example.

     The natural evolution of microwave radiometry is the continual push into
the millimeter-wavelengths (300 GHz and above) as receiver components become
available at these frequencies; see WILSON (1983) and FUNG et al. (1982). The
advantages at these shorter wavelengths are more abundant microwave lines and
generally stronger lines as well. The disadvantage is the higher tropospheric
opacities especially above about 100 GHz due almost entirely to water vapor (see
Figure 3). Submillimeter-wavelength radiometry from the ground will require
dry, preferably high altitude sites (similar to the needs of optical and
infrared astronomy).

     Finally, the almost complete computer control (automation) of microwave
facilities in the next few years and their round-the-clock use in middle
atmosphere observations will produce an information stream second only to
satellite experiments.

CONCLUDING REMARKS

     In the past decade we have witnessed the growth and acceptance of ground-
based microwave radiometry as a most useful measurement technique. During the
remainder of this decade we will observe continued expansion of this technique
to new field sites, to new continents, to dedicated and long-term experiments,
to cooperative-collaborative groupings of composition and temperature sensors,
to improved receiver technology, to new species, and to observing atmospheric
lines at ever higher frequencies (at submillimeter-wavelengths).
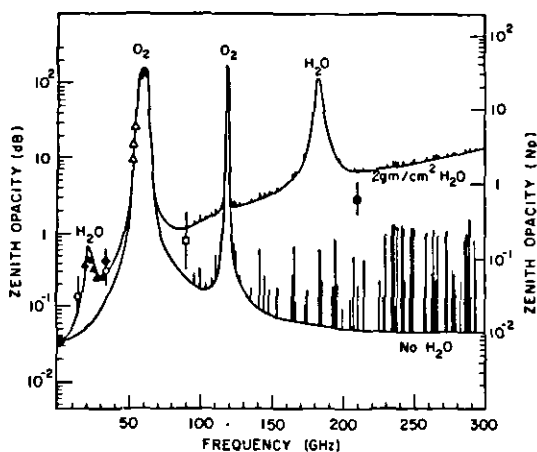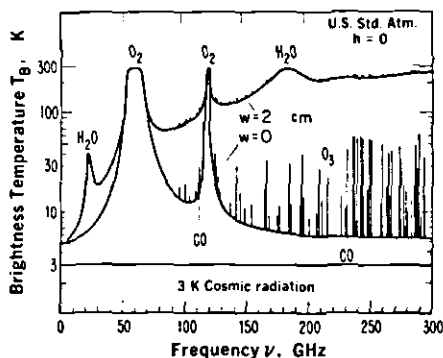
Figure 3. Atmospheric microwave spectrum below
300 GHz; (a) Opacity in the zenith, with and
without 2 cm (2 gm/cm^2) precipitable water
(overburden); (b) Brightness temperature viewed
at the zenith, with and without 2 cm
precipitable water (WATERS, 1976; LIEBE, 1980).

REFERENCES

Anderson, J. G., H. J. Grassl, R. E. Shetter and J. J. Margitan (1977), Strato-
   spheric free chlorine measured by balloon-borne in situ resonance
   fluorescence, J. Geophys. Res., 85, 2869-2878.
Barrett, A. H. and V. K. Chung (1962), A method for the determination of high
   altitude water vapor abundance from ground-based microwave observations,
   J. Geophys. Res., 67, 4259-4266.
Berg, N. J., M. W. Casseday and I. J. Abramovitz (1982), Acousto-optic
   processing increases EW capabilities, Microwave Systems News, 12,
   106-118.

54

Berg, N. J. and J. N. Lee (Editors) (1982), Acousto-Optical Signal Processing, Marcel Decker, Inc., New York.

Bevilacqua, R. M. (1982), An observational study of water vapor in the mid-latitude mesosphere using ground-based microwave techniques, Ph.D. Dissertation in Meteorology, The Pennsylvania State University.

Bevilacqua, R. M., P. R. Schwartz, J. M. Bologna, D. L. Thacker, J. J. Olivero and C. J. Gibbins (1983), An observational study of water vapor in the mid-latitude mesosphere using ground-based microwave techniques, J. Geophys. Res., 88, 8523-8534.

Caton, W. M., G. G. Manella, P. M. Kalaghan, A. E. Barrington and H. I. Ewen (1968), Radio measurement of the atmospheric ozone transition at 101.7 GHz, Astrophys. J., 151, L153-L156.

Clancy, R. T., D. O. Muhlman and M. Allen (1984), Seasonal variability of CO in the terrestrial mesopshere, submitted for publication.

Clancy, R. T., D. O. Muhlman and G. L. Berge (1982), Microwave spectra of terrestrial mesospheric CO, J. Geophys. Res., 87, 5009-5014.

Cooper, B. F. C. (1976), Autocorrelation spectrometers, in Methods of Experimental Physics, vol. 12, Astrophysics, part B, Radio Telescopes, edited by M. L. Meeks, 180-198, Academic Press, New York.

Croom, D. L. (1965), Stratospheric thermal emission and absorption near the 22.235 Gc/s (1.35 cm) rotational line of water vapor, J. Atmos. Terr. Phys., 27, 217-233.

Croskey, C. L., J. D. Carlson, L. C. Hale, J. J. Olivero, J. J. Tsou and R. G. Joiner (1983), A comparison of ionospheric propagation to microwave measurements of mesospheric water vapor, National Radio Science Meeting (URSI), Boulder.

de La Noe, J., A. Baudry, M. Perault, P. Dierich, N. Monnanteuil and J. M. Colmont (1983), Measurements of the vertical distribution of ozone by ground-based microwave techniques at the Bordeaux Observatory during the June 1981 Intercomparison Campaign, Planet. Space Sci., 31, 737-741.

de Zafra, R. L., A. Parrish, P. M. Solomon and J. W. Barrett (1984), A measurement of stratospheric HO_2 by ground-based mm-wave spectroscopy, J. Geophys. Res., 89, 1321-1326.

Froidevaux, L. (1983), Photochemical modeling in the Earth's stratosphere, Ph.D. Dissertation in Planetary Sciences, California Inst. of Technology.

Fung, S. W., M. T. Lebenbaum and F. P. Parini (1982), Mm-wave receiver front-ends reach out to space, Microwave Systems News, 12, 103-110.

Gibbins, C. J., P. R. Schwartz, D. L. Thacker and R. M. Bevilacqua (1982), The variability of mesospheric water vapor, Geophys. Res. Lett., 9, 131-134.

Goldsmith, P. F., M. M. Litvak, R. L. Plambeck and D. R. W. Williams (1979), Carbon monoxide mixing ratio in the mesosphere derived from ground-based microwave measurements, J. Geophys. Res., 84, 416-418.

Kunzi, K. F. and E. R. Carlson (1982), Atmospheric CO volume mixing ratio profiles determined from ground-based measurements of the J=1-0 and J=2-1 emission lines, J. Geophys. Res., 87, 7235-7241.

Liebe, H. J. (1980), Atmospheric water vapor: A nemesis for millimeter wave propagation, in Atmospheric Water Vapor, edited by A. Deepak, T. D. Wilkerson and L. H. Ruhnke, Academic Press, New York.

Lobsiger, E., K. F. Kunzi and H. U. Dutsch (1983), Accuracy of atmospheric ozone profiles measured with a ground-based microwave sensor, Symposium on Middle Atmosphere Sciences, International Union of Geodesy and Geophysics, XVIII General Assembly, Hamburg.

Mumma, M. J., J. D. Rogers, T. Kostiuk, D. Deming, J. J. Hillman and D. Zipoy (1983), Is there any ClO in the stratosphere?, Science, 221, 286-288.

Parrish, A., R. L. deZafra, P. M. Solomon, J. W. Barrett and E. R. Carlson (1981), Chlorine oxide in the stratospheric ozone layer: Ground-based detection and measurement, Science, 211, 1158-1161.

Penfield, H. (1976), Multichannel-filter spectrometers, in Methods of Experimental Physics, vol. 12, Astrophysics, part B, Radio Telescopes, edited by M. L. Meeks, 266-279, Academic Press, New York.

Penfield, H., M. M. Litvak, C. A. Gottlieb and A. E. Lilley (1976), Mesospheric ozone measured from ground-based millimeter-wave observations, J. Geophys. Res., 81, 6115-6120.

Radrurd, H. E., M. M. Litvak, C. A. Gottlieb, E. W. Gottlieb, S. K. Rosenthal and A. E. Lilley (1977), Mesospheric water vapor measured from ground-based microwave observations, J. Geophys. Res., 82, 472-478.

Schwartz, P. R., C. L. Croskey, R. M. Bevilacqua and J. J. Olivero (1983), Microwave spectroscopy of H_2O in the stratosphere and mesosphere, Nature, 305, 294-295.

Shimabukuro, F. I., P. L. Smith and W. J. Wilson (1975), Estimation of ozone distribution from millimeter-wavelength absorption measurements, J. Geophys. Res., 80, 2957-2959.

Shimabukuro, F. I., P. L. Smith and W. J. Wilson (1977), Estimation of the daytime and nighttime distribution of atmospheric ozone from ground-based millimeter-wavelength measurements, J. Appl. Meteorol., 16, 929-934.

Shimabukuro, F. I. and W. J. Wilson (1973), Observations of atmospheric ozone at 110.836 GHz, J. Geophys. Res., 78, 6136-6139.

Solomon, P. M., R. de Zafra, A. Parrish and J. W. Barrett (1983), Observation of diurnal variation of stratospheric chlorine oxide: A critical test of chlorine chemistry in the ozone layer, Atmospheric Spectroscopy Symposium, Oxford, England.

Thacker, D. L., C. L. Gibbins, P. R. Schwartz and R. M. Bevilacqua (1981), Ground-based microwave observations of mesospheric H_2O in January, April, July and September 1980, Geophys. Res. Lett., 8, 1059-1062.

Twomey, S. (1977), Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Sensing, Elsevier, New York.

Ulaby, F. T., R. K. Moore and A. K. Fung (1981), Microwave remote sensing: Active and passive, Vol. 1, Microwave Remoting Sensing Fundamentals and Radiometry, pp. 456, Addison-Wesley Pub. Co., Reading, MA.

Waters, J. W. (1973), Ground-based measurement of millimeter-wavelength emission by upper stratospheric O_2, Nature, 242, 506-508.

Waters, J. W. (1976), Absorption and emission by atmospheric gases, in Methods of Experimental Physics, vol. 12, Astrophysics, part B, Radio Telescopes, edited by M. L. Meeks, 142-176, Academic, New York.

Waters, J. W., W. J. Wilson and F. I. Shimabukuro (1976), Microwave measurement of mesospheric carbon monoxide, Science, 191, 1174-1175.

Wilson, W. J. (1983), Submillimeter-wave receivers - A status report, IEEE Trans. Microwave Theor. Tech., MTT-31, 873-878.

Wilson, W. J. and P. R. Schwartz (1981), Diurnal variations of mesospheric ozone using millimeter-wave measurements, J. Geophys. Res., 86, 7385-7388.

5. LIDAR SYSTEMS FOR AEROSOL STUDIES: AN OUTLINE

G. Fiocco

Dipartimento di Fisica
Citta Universitaria
00185 Roma, Italy

ABSTRACT

An introductory survey of the problems encountered in the design and operation of a ground based lidar system and its application to atmospheric aerosol studies in the height region 10-100 km.

GENERALITIES ON LIDAR SYSTEMS AND THEIR APPLICATIONS

The lidar (acronym for LIght Detection and Ranging), sometimes called optical or laser radar, is an active remote sensing device based on the radar principle which has found wide applications in atmospheric studies. In what follows,its use for detecting the presence of aerosols and determining their physical properties in the height interval of interest to the Middle Atmosphere Program, is outlined.

In principle, a lidar provides measurements of the optical backscattering cross section of air as a function of range and, possibly, wavelength. This information may be subsequently interpreted to obtain profiles of the aerosol concentration, size distribution, refractive index, scattering, absorption and extinction cross sections, and shape. The feasibility of aerosol velocity measurements has also been demonstrated: a lidar may thus become a remote sensor of wind motions, utilizing aerosols as tracers.

Ground based lidars have been operated for about two decades (for reviews see e.g., FIOCCO, 1971; COLLIS and RUSSELL, 1976); airborne systems have also been implemented and studies for deploying a lidar in the Space Shuttle have been sponsored by the European Space Agency (see e.g., FIOCCO et al., 1976) and by the National Aeronautics and Space Administration (see e.g., RUSSELL et al., 1982a,b).

Aerosols play a very important role in the radiative budget of the planet and are sensitive indicators of the atmospheric structure, composition and dynamics. For the sake of studies related to their presence and to the utilization of lidars, the height interval 10-100 km can be further subdivided into regions - the troposphere, the lower and middle stratosphere, the meso- sphere, the mesopause - where the questions related to their origin, existence and properties have a distinct character.

The region around the tropopause is characterized by temperature minima which favor the formation of stratified, optically thin, ice clouds. This transition region may have a different character in the tropics, due to the upgoing branch of the Hadley cell and the related action of the Brewer-Dobson mechanism, compared to higher latitudes. Above it, a progressive change with height of the aerosol properties, in particular their water content, takes place. At all latitudes the lower stratosphere is characterized by the irregular presence of aerosols, composed of mixtures of water and sulfuric acid, which are the result of in situ gas-to-particle conversion processes; once formed, their lifetime is long, of the order of one year. The concentration of stratospheric aerosols is strongly dependent on volcanic activity, with measurable effects on the radiation field and atmospheric dynamics. Of quite different origin, in the high-latitude stratosphere, nacreous (mother of pearl) clouds are occasionally observed. Noctilucent clouds are often visible in the

high-latitude summer mesopause. Both noctilucent and nacreous clouds are
currently believed to be composed of ice. Mesospheric stratifications have
sometimes been detected but their nature is elusive.

Lidars have been successfully employed to observe many of these phenomena,
after the early observations of the stratospheric aerosol layers following the
Mt. Agung eruption of 1963 (FIOCCO and GRAMS, 1964). A trace with time of
developments and results can be found in the proceedings of the regularly held
International Laser Radar Conferences. Lidar systems have presently a high
degree of reliability and useful measurements can be carried out on a continuous
basis with relatively simple, easy to operate and maintain instruments. A
number of lidar systems sufficiently sensitive to carry out measurements above
the tropopause are presently operational; for the most part they are located in
the mid-latitude Northern Hemisphere (see Table 1).

Table 1. Lidar Stations Presently Active.

Aberystwyth	52.4°N	4.1°E
Haute Provence	44°N	5°E
Garmisch-Partenkirchen	47.5°N	11°E
Frascati	41.49°N	12.40°E
L'Aquila	42°N	13°E
Fukuoka	33.7°N	130.4°E
Nagoya	35°N	137°E
Mauna Loa	19.5°N	130.35°E
Boulder	40°N	105.4°W
Hampton	37.1°N	76.3°W
San Jose dos Campos	23°S	46°W

Observations at stations listed in the table (ADRIANI et al., 1983;
CLEMESHA and SIMONICH, 1983; COULSON et al., 1982; D'ALTORIO and VISCONTI, 1983;
HIRONO and SHIBATA, 1983; IWASAKA et al., 1983; LEFRERE et al., 1983; REITER et
al., 1983; etc.) are making it possible, for example, to follow in detail the
evolution of the stratospheric aerosol clouds after the recent eruption of the
volcano El-Chichon (28 March – 4 April 1982).

It should be pointed out that while the aerosol concentration has a high
spatial and temporal variation, existing lidar stations are unevenly distributed
and often operated in a discontinuous fashion. The setting up of an inter-
national network, with appropriate support, should be considered.

THE LIDAR EQUATION

In general a lidar consists of a source of optical radiation that is
collimated into a reasonably narrow beam and pulsed for purposes of range
determination and discrimination against short range unwanted echoes, a
receiving telescope, means for spectral filtering and conversion of the optical
signals into electrical signals, suitable electronics for display, analysis and
storage.

In a vertically pointing, monostatic configuration, where the axes of the
transmitted and received beams are essentially coaxial, for a short pulse of
transmitted energy W_t the power P_r received from scatterers located at a
distance z, on the hypothesis of single scattering, can be expressed by the
simple equation:

$$P_r = W_t \ (c/2) \ \beta(z) \ z^{-2} \ A \ K \ \exp(-2\int_0^z \alpha(z')dz') \tag{1}$$

c is the speed of light, β is the backscattering (differential) cross section per unit volume of air, A is the collecting area of the antenna, K is a coefficient that takes into account instrumental effects such as beam overlapping and efficiencies, α is the atmospheric extinction coefficient. Backscattering can be expressed as the sum of two components, β_M and β_A, respectively due to the gaseous and the aerosolic constituents of air:

$$\beta = \beta_M + \beta_A \tag{2}$$

Backscattering by the gaseous constituents can be expressed as

$$\beta_M = N_M \ (dC_{scM}/d\Omega)_{180°} \tag{3}$$

where N_M is the molecular number density of air, and the average differential backscattering cross section of one air molecule is given by:

$$(dC_{scM}/d\Omega)_{180°} = 5.45[\lambda(\mu m)/0.55]^{-409} \ 10^{-28} \ cm^2 \ sr^{-1} \tag{4}$$

The extinction coefficient also takes into account both the gaseous and the particulate constituents and can be schematically written as:

$$\alpha = \alpha_M + \alpha_A \tag{5}$$

For treatments of the topic of light transmission through the atmosphere see e.g., ZUEV (1976), and the various developments of the Lowtran code (KNEYZIS et al., 1980).

Extinction and backscattering are in principle related. For the gaseous constituents, neglecting absorption and assuming therefore that α_M is due to scattering only, the following simple relation holds:

$$\alpha_M/\beta_M = (4\pi/1.5) \ sr = 8.3776 \ sr \tag{6}$$

In general vertical profiles of the number density of air for the lower stratosphere can be retrieved with sufficient accuracy through the weather services' data; thus the molecular contributions to backscattering and extinction can be easily estimated. Lacking such data, reference can be made to a standard atmosphere with some reduction in accuracy. The scattering and extinction properties of aerosols are more involved, and will be considered with some detail in a later section.

The results of the aerosol lidar measurements are generally presented in terms of the scattering ratio:

$$R = \beta_M + \beta_A / \beta_M \tag{7}$$

To obtain R from the echo intensity involves separating the aerosol from the molecular echoes having taken care of the extinction.

In experiments where only the relative intensity of the light received is recorded as a function of the echo delay-time, it is in general difficult to separate the contributions of the molecules from those of the aerosol. If the optical radar were absolutely calibrated and the atmospheric transmission and molecular composition were known, it would be possible, by knowing a priori the contribution of air molecules, to attribute the excess signal to the aerosol. This is in reality not feasible. It is standard practice to deduce vertical profiles of the aerosol backscattering, by assuming that such quantity is known (or negligible) in a certain height interval. In this respect, the generally layered nature of the aerosol echoes is helpful.

From eqn. 1 it appears that the intensity of the received signal depends primarily on the backscattering, and secondarily on the extinction. For light aerosol loads the effect of the aerosol extinction on the intensity of the echoes is minor but it becomes important and should not be neglected when the loads are large, e.g., in the aftermath of volcanic eruptions.

For the sake of obtaining R, the extinction is however only a correction term and as such may not need to be known to a great accuracy. On the other hand the extinction is a quantity of greater geophysical relevance than backscattering. For this purpose it can be inferred through successive approximations from a knowledge of R, with the aid of model calculations and of any additional information on the aerosol characteristics obtainable by other means.

For spherical aerosol, given the size distribution and the refractive index, the ratio of extinction to backscattering can be calculated: for an early approach to this problem see MCCORMICK et al. (1968). If such characteristics are not known, it is possible to proceed in a semiempirical way by comparing the lidar backscattering results with the extinction measured by other means. On this basis SWISSLER et al. (1983) have adopted the following relationship:

$$\alpha_A/\beta_A = 42.8 \pm 7.4 \text{ sr}, \ (\lambda = 0.6943 \ \mu m) \tag{8}$$

In principle it should be possible to infer the extinction directly by lidar techniques, in a consistent way, by diversifying the experiment and achieve a good level of accuracy and precision. It has been proposed to carry out observations at slant angles (see e.g., SPINHIRNE et al., 1980), but the additional loss of signal due to the increased range and lack of horizontal homogeneity may present problems.

A few words should be spent on the precision of the measurements. Lidar data exhibit pulse-to-pulse fluctuations attributable in part to the variability of the aerosol concentration and in part to other sources of noise. The fluctuations can be reduced either by integrating successive pulses or by reducing the vertical resolution.

Figure 1 is an example of application of the procedures previously outlined. Solid lines indicate the intensity of the backscattered echoes as a function of height, for three different amounts of integration. The data, taken at Frascati (ADRIANI et al., 1983), show stratifications related to the El Chichon eruption. The profiles have been adjusted for the effects of extinction. Also shown are the estimated returns from a molecular atmosphere. The unit of relative intensity is equivalent to one step in the analog-to-digital (A/D) conversion.

Figure 2 shows the scattering ratios thus obtained. As in Figure 1 the three curves refer to different integration amounts, namely single shot, 10 shots (time interval 24 s), 100 shots (time interval 240 s). The amount of correlation existing in the echoes from differently spaced shots provides a measure of the noise level of the measurements and of the variability with time of the aerosol cross section.

With reference to Figures 1 and 2, since the received echoes are sampled at the rate of 0.5 μs, and the laser pulse has a duration of the order of 10 μs, the resulting vertical resolution of the profiles is 75 m. Because of the high sensitivity of the system, this resolution can be effectively preserved without recourse to spatial integration and that, at least in the lower part of the record, only a modest amount of time integration is sufficient to smooth out fluctuations that can be classified as noise.
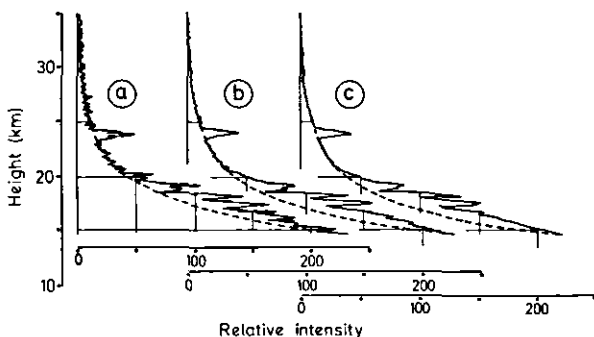
60

Figure 1. Lidar backscattering profiles (solid
lines) and expected molecular returns (dotted
lines), for various degrees of integration:
(a) single shot; (b) 10 shots (24s); (c) 100
shots (240s). The measurements taken on
9 June 1982 indicate the presence of strong
aerosol layers, consequence of the El Chichon
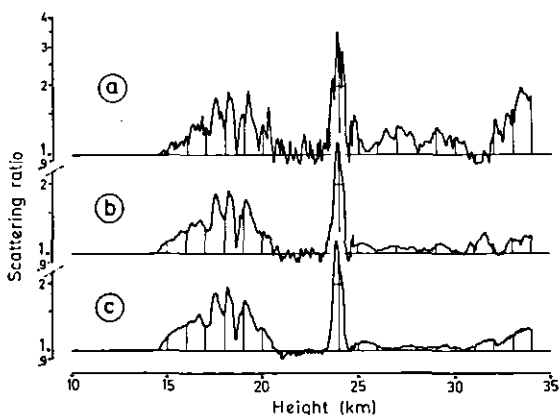eruption (from ADRIANI et al., 1983).

Figure 2. Profiles of the scattering
ratio R obtained from the curves in
Figure 1 (from ADRIANI et al., 1983).

It should be mentioned that a different and more satisfactory way to
separate the molecular from the aerosol contributions, which however involves
a substantial increase in technical complexity, is based on the spectral
analysis of the echoes (FIOCCO et al., 1971). Doppler broadening which affects
molecular echoes is essentially due to random thermal motions and is larger than
the broadening associated with the Brownian motions of the more massive aerosols
(see Figure 3); from the different spectral signatures of their returns, the two
contributions can be separated. It is expected that spectral analysis
techniques, based on a better utilization of the lasers' coherence, should
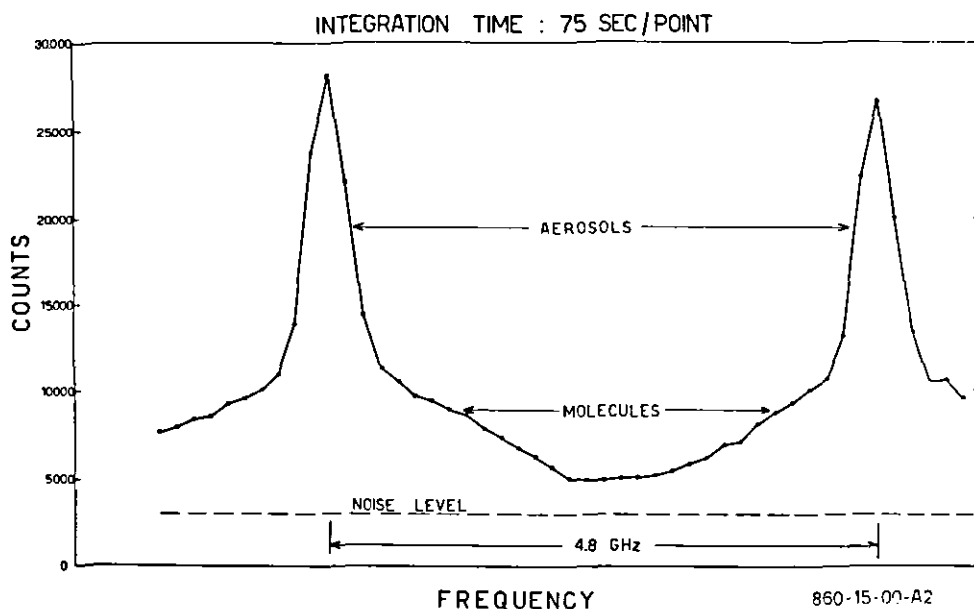provide a substantial improvement in the amount of information obtained.

INTEGRATION TIME : 75 SEC/POINT

Figure 3. Spectra of the backscattered signal;
the molecular contributions and the aerosol
contributions can be separated.

MISCELLANEOUS TECHNICAL CONSIDERATIONS

Although bistatic atmospheric scattering experiments using searchlights had
been suggested by Synge in 1930 and later carried out, it is only with the
invention of the laser that much interesting work has been made possible. We
shall not discuss the various types of lasers that can be used: the important
characteristics are the average and peak power output; the spatial coherence and
divergence of the beam; the output wavelength, spectral characteristics, mode
structure and stability; type of operation, e.g., pulsed or continuous wave;
suitability for modulation and for harmonic generation; tunability; efficiency,
etc.

The first lidars utilised ruby lasers as sources; subsequently several
different laser families became available, e.g., dye, gas, Nd, excimer, etc.
We shall not consider these sources in detail, nor shall we consider so-called
"coherent" lidars whereby detection is accomplished by heterodyne. As an
example of a powerful lidar operating with conventional detection schemes we
mention the system presently working at Frascati: it incorporates a Nd YAG
laser equipped with a frequency doubler and capable of firing pulses
simultaneously at two wavelengths, namely 1.06 μm and 0.53 μm. The output
energy per pulse is approximately 0.5 J at either wavelength with a firing rate
of 10 Hz.

A beam expander is often inserted in the transmitting optics of a lidar
system to avoid damage to optical surfaces. Collimating the outgoing energy in
a low-divergence beam is also often a desirable measure to reduce the field of
view of the receiver for the sake of limiting the sky background. Another
element of the transmitter optics may be a shutter introduced to positively
preclude feedthroughs in the receiver.

The light collector or antenna is basically a telescope: possibly of large aperture and quite reduced resolution, generally pointing to the zenith. Different optical configurations have been used, the Cassegrain having been often favored, in view of its compactness. The main consideration for these applications is the size of the primary collecting area.

In general, the quality of the optics is dictated by the maximum field of view which can be tolerated in order to reduce the sky background and feed the received optical signals into a detector or a spectral analysis system. Background light does not generally represent a major problem for nighttime operation, with the exception of the high latitude summer. Daytime operation may, on the other hand, introduce substantial difficulties.

Lidars useful for observing the stratosphere have been built around *astronomical telescopes with diameters* ranging from approximately 0.4 m to over 1.0 m. Since weight and price of single-element glass mirrors increase considerably with size, alternate solutions have been sought, utilizing metallic mirrors, or assemblies of smaller mirrors. In particular, the mosaics assembled in Kingston (KENT et al., 1971; presented not in use) and in Frascati should be mentioned. It is worth pointing out, however, that much interesting work can be carried out with relatively small optics.

The Frascati mosaic consists of 36 hexagonally-shaped, 0.2 m apothem, 16 m radius-of-curvature, spherical mirrors arranged and individually adjusted to approximate a primary paraboloidal surface of about 3 m diameter and a focal length of 8 m. The total collecting area is thus 5.0 m and the total angular resolution is 1.0 mrad. Although the detectors are presently placed at the primary focus, the system was designed and built also to operate as a Cassegrain by introducing a secondary mirror, with a resulting focal length of approximately 48 m.

Questions related to the operation of the photodetectors may be quite involved but, in the end, the options are limited. In choosing a detector, which in general is a photomultiplier, one should bear in mind such often variable and imperfectly known characteristics as quantum efficiency, gain, bandwidth, saturation, dynamic range, dark current, and fatigue.

In the Frascati system a dichroic mirror is used to separate the signals received at the two wavelengths; these are subsequently filtered and detected with the aid of two photomultipliers, respectively type EMI 9684 (S1) and EMI 9558 (S20). The S1 tube is cooled.

Cooling a photomultiplier generally leads to a reduction of its noisiness or dark current: the effect is different from type to type, the causes of noise being not only a function of temperature. Sometimes the photomultiplier noise may be neglected in comparison with the sky background. Use of small cathodes is another approach to reducing the dark current; in this respect magnetic fields have been used to limit the effective portion of a large cathode. Cooling below, say, -10°C may not lead to further noise reduction since other internal causes may then prevail; large reduction of the noise is progressively obtained after keeping the tube in darkness with the high voltage applied.

The introduction of fast shutters prevents the detector from getting flooded by unwanted echoes, reducing transients and fatigue. Shutters may consist of a wheel driven at high speed (e.g., 12000 ro 24000 rpm) by a small synchronous motor: in our experiences, because of the limited mechanical power available, the dimension of the wheel cannot exceed a maximum diameter of about 10 cm unless it operates in an enclosure which is either evacuated or filled with a light gas.

Filtering and spectral analysis is a vast subject and will not be considered here, although, as already pointed out, useful developments are foreseeable in that direction. For the sake of measurements of current interest bandwidths down to a few Angstroms can be obtained with commercially available filters: stability and fluorescence may pose problems.

The wide dynamic range of the incoming signal poses severe limitations to the performance of the detecting and amplifying system. Critical is the ability to recover from the transients due to strong, short-range echoes. Particularly critical is the behavior of diode circuits sometimes placed in the circuitry for protection. Shutters and gating circuits may have to be introduced to reduce such problems. Gain switching with distance is an artifice used to extend the dynamic range of the amplification system.

The dynamic range may pose a serious limitation in the A/D conversion of the signals for further storage and analysis. Older transient analyzers have a dynamic range of 8 bits, more modern ones have 11 bits or better. If the system gain is adjusted to make observations possible at a lower range of, say, 10 km, the limited dynamic range will provide insufficient resolution above a range of approximately 30 km for an 8 bit converter. This effect is equivalent to the introduction of an additional source of noise which cannot be easily reduced by extended statistical analysis.

If the gain and bandwidth characteristics of the detector-amplifying system permit, the quantized nature of the photoelectric curent can be preserved, and the intensity of the received signal can be measured by counting the number of photoelectrons per range interval. When, however the radiant fluxes are too high, the individual nature of the amplified photoelectrons in the stream tends to disappear, and several photons may merge into a single spike, with a loss in the counting rate. The counting technique is therefore particularly advantageous for high altitude observations, when the backscattered fluxes are low. The counting technique lends itself easily to signal integration.

CONSIDERATIONS ON THE RETRIEVAL OF ATMOSPHERIC AEROSOL PROPERTIES

The analysis of lidar data poses the inverse problem of inferring the physical characteristics of the aerosols from their volume backscattering characteristics.

For a general reference to the theory of scattering by aerosols see KERKER (1969). Radiative transfer and inversion problems with specific reference to the use of lidar have been approached by several authors (see e.g., LIOU and SCHOTLAND, 1971; REAGAN et al., 1977; KLETT, 1981; the extended monograph of ZUEV and NAATS, 1982, which contains many references to Russian publications). CADLE and GRAMS (1975), and TURCO et al. (1982), have provided useful reviews of stratospheric aerosol properties.

With regard to the optical properties, analytical solutions exist for spherical particles (Mie formula) and for nonspherical particles with modest amounts of asymmetry. Individual scattering cross sections of spheres are defined, given the complex refractive index $\tilde{n} = \tilde{n}' - j\tilde{n}''$, and the radius r or the normalised Mie size parameter $2\pi r/\lambda$.

The differential backscattering cross section of a particle of radius r can be expressed as

$$(dC_{scA}/d\Omega)_{180°} = \pi r^2 \ Q_{sc} \tilde{P}(180°) \tag{9}$$

where Q_{sc} is the scattering efficiency and $\tilde{P}(\theta)$ is the phase function.

Also used sometimes is the radar cross section whose value is 4π times the backscattering differential cross section. Curves in Figure 4 show the radar cross section as a function of Mie size for spheres with a real part of the refractive index $\hat{n} = 1.55$ and different values of the imaginary part.
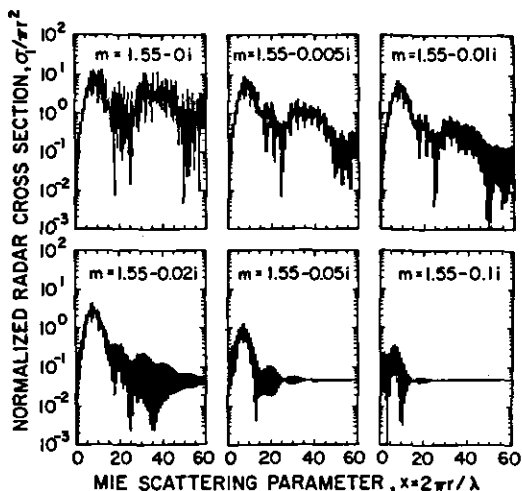

Figure 4. Aerosol normalized cross section for various values of the refractive index, as a function of the Mie parameter (from CADLE and GRAMS, 1975).

In reality particles are distributed with a variety of radii and the volume cross sections are obtained by integration:

$$\beta_A = \int_{r_{min}}^{r_{max}} \pi r^2 \, Q_{sc} \, \hat{P}(180°) \, (dN/dr) \, dr \tag{10}$$

where $N(r)$ is the number of particles with radius larger than r.

Size distributions as occur in nature have to be approximated for the sake of performing model calculations, although doubts may sometimes arise as to the validity of oversimplified approaches.

Different types of analytic functions have been adopted to describe the stratospheric aerosol size distributions, e.g., Junge, log normal, etc. Widely used is the so-called zero order logarithmic distribution function (ZOLD):

$$dN/dr = C \exp\left[-(\ln r - \ln r_m)^2 / 2 \ln^2\sigma\right] \tag{11}$$

where r_m and σ define respectively the model radius and the width of the distribution. This function has been used to fit available stratospheric aerosol distribution data by TOON and POLLACK (1976) with a choice of $\sigma = 2.0$, $r = 0.035$ μm. Figure 5 shows values of the averaged backscattering cross section vs. model radius for a ZOLD with $\sigma = 1.8$.

Hofmann and Rosen, through the use of balloon-borne dust sondes, have measured size distributions in the stratosphere, indicating (HOFMANN and ROSEN,
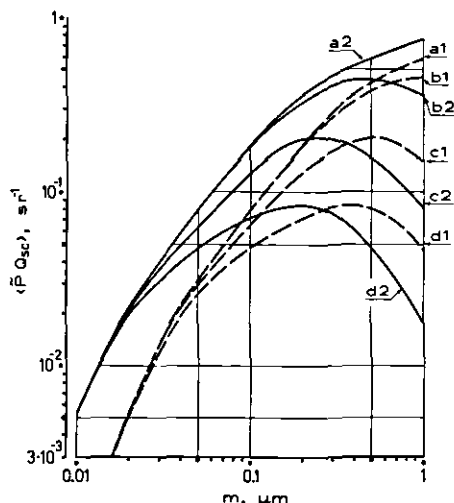
Figure 5. Volume backscattering for spheres with $\tilde{n}' = 1.43$, and various values of $\tilde{n}''$, for a ZOLD ($\sigma = 1.8$), as a function of the model radius; (a) $\tilde{n}'' = 0.000$, (b) $\tilde{n}'' = 0.001$, (c) $\tilde{n}'' = 0.010$, (d) $\tilde{n}'' = 0.032$. Suffixes 1 and 2 refer to $\lambda_1 = \lambda 0.6$ μm and $\lambda_2 = 0.532$ μm.

1982) that a better match would be obtained by using bimodal distributions, and later carrying out comparison with lidar returns (HOFMANN et al., 1983). Data subsequently obtained, shortly after the El Chichon eruption, showed the difficulty of fitting standard distributions (HOFMANN and ROSEN, 1983). In a recent study of the temporal evolution of the cloud they obtain log normal fits, after approximately 200 days, with mean radii around 0.15 m in the 17-20 km region and around 0.27 μm in the 20-25 km region, and around 0.22 μm after approximately 400 days (HOFMANN and ROSEN, 1984).

Indications about the aerosol characteristics can be obtained by carrying out lidar measurements at more than one wavelength (see e.g., IWASAKA and ISONO, 1977; HIRONO and SHIBATA, 1983; and others). The diagrams in Figures 6, 7 and 8 show the ratio of the cross sections for a ZOLD, computed at two wavelengths, respectively $\lambda_1 = 1.06$ μm, and $\lambda_2 = 0.53$ μm, as a function of the model radius.

The curves in Figure 6 show the effect of changing the real part of the refractive index for non-absorbing particles from the value $\tilde{n}' = 1.33$, typical of H_2O, to higher values, for a width $\sigma = 1.8$. Changes of $\tilde{n}'$ are expected in the lower stratosphere due to a change with altitude of the H_2SO_4 concentration in the solution.

The following figure refers to various values of the width of the distribution, and a fixed value of the refractive index ($\tilde{n} = 1.43 - 0.000$), characteristic of 75% H_2SO_4/H_2O mixtures. The ratios thus obtained are large for small values of the model ratio and tend to approach unity for large values.

Finally the curves in Figure 8 show the effect of changing the absorptivity by assuming different values of the imaginary part $\tilde{n}''$: large variations are obtained in the range of model radii between 0.1 and 1.0 μm.

From these graphs it appears that the controlling factors are not only the model ratio and the width of the distribution, as idealized parameters as these can be, but also the refractive index and particularly its imaginary part. It should be considered that these values control the extinction and above all the heating rate induced by the aerosol in the stratosphere and have therefore a definite significance on the climatic impact of large aerosol loads.
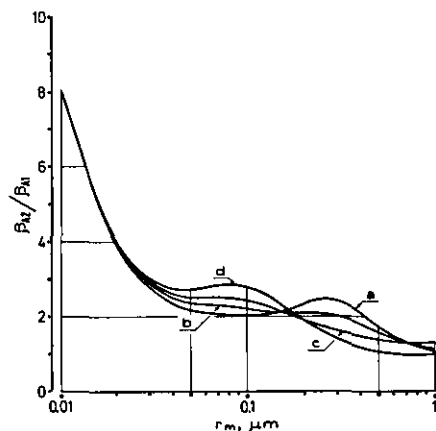
Figure 6. Ratio of backscattering coefficients at $\lambda_1 = 1.06$ μm and $\lambda_2 = 0.53$ μm for spheres distributed according to a ZOLD, as a function of the model radius. Different curves refer to various values of $\tilde{n}' = \tilde{n}'_1 = \tilde{n}'_2$, and $\tilde{n}''_1$ and $\tilde{n}''_2 = 0.000$: (a) $\tilde{n}' = 1.33$; (b) $\tilde{n}' = 1.39$; (c) $\tilde{n}' = 1.43$; (d) $\tilde{n}' = 1.49$.
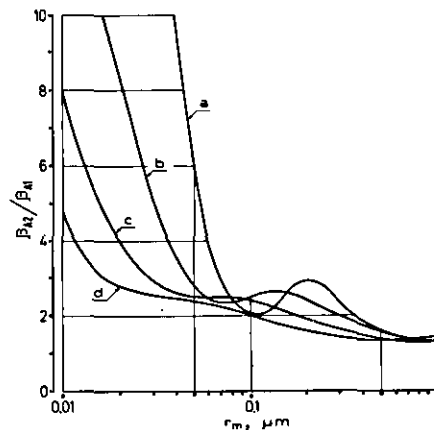
Figure 7. Same as Figure 6. Different curves refer to various values of the width of the distribution and a fixed value of refractive index ($\tilde{n} = 1.43 - j0.000$); (a) $\sigma = 1.4$; (b) $\sigma = 1.6$; (c) $\sigma = 1.8$; (d) $\sigma = 2.0$.
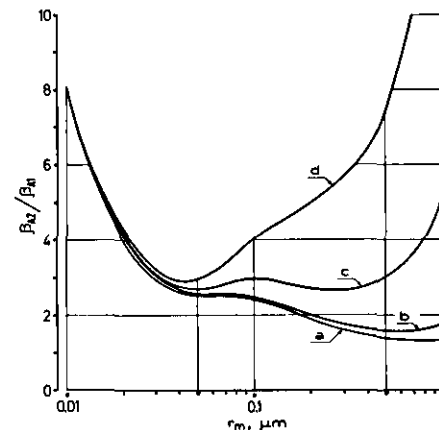
Figure 8. Same as Figure 6, but for different values of the imaginary part of the refractive index, for $\tilde{n}'_1 = \tilde{n}'_2 = 1.43$, $n''_2 = 0$, and a fixed value of the width ($\sigma = 1.8$): (a) $n''_1 = 0.000$; (b) $\tilde{n}''_1 = 0.001$; (c) $\tilde{n}''_1 = 0.01$; $\tilde{n}''_1 = 0.032$.

The author apologizes to the many colleagues whose contributions have not been adequately references. Thanks are due to A. Mugnai and R. Ligi for the calculations on which Figure 5 through 8 are based.

REFERENCES

Adriani, A., F. Congeduti, G. Fiocco and G. P. Gobbi (1983), One-year lidar observations of the stratospheric aerosol at Frascati, March 1982 - March 1983, Geophys. Res. Lett.

Cadle, R. D. and G. W. Grams (1975), Stratospheric aerosol particles and their optical properties, Rev. Geophys. Space Phys., 13, 475-501.

Collis, R. T. H. and P. B. Russell (1976), Lidar measurement of particles and gases by elastic backscattering and differential absorption, in Laser Monitoring of the Atmosphere, E. D. Hinkley, editor, Springer-Verlag.

Coulson, K., T. E. Defoor and J. DeLuisi (1982), Lidar and optical polarization measurements of stratospheric cloud in Hawaii, EOS Trans. AGU, 63, 897.

D'Altorio, A. and G. Visconti (1983), Lidar observations of dust layers' transience in the stratosphere following the El Chichon volcanic eruption, Geophys. Res. Lett., 10, 27-30.

Fiocco, G. (1971), Use of optical radars for atmospheric studies, in Physics of the Upper Atmosphere, F. Verniani, editor, Editrice Compositori, Bologna

Fiocco, G., G. Benedetti-Michelangeli, K. Maischberger and E. Madonna (1971), Measurement of temperature and aerosol-to-molecule ratio in the troposphere by optical radar, Nature Phys. Sci., 229, 78-79.

Fiocco, G. and G. Grams (1964), Observations of the aerosol layer at 20 km by optical radar, J. Atmos. Sci., 21, 323-324.

Fiocco, G., W. Renger and L. Thomas (1976), Spacelab-borne lidar-scientific objectives, in Atmospheric Physics from Spacelab, J. J. Burger et al., editors, Reidel.

Hirono, M. and T. Shibata (1983), Enormous increase of stratospheric aerosols over Fukuoka due to volcanic eruption of El Chichon in 1982, Geophys. Res. Lett., 10, 152-154.

Hofmann, D. J. and J. M. Rosen (1982), Balloon-borne observations of stratospheric aerosol and condensation nuclei during the year following the Mt. St. Helen's eruption, J. Geophys. Res., 87, 11039-11061.

Hofmann, D. J. and J. M. Rosen (1983), Stratospheric sulfuric acid fraction and mass estimate for the 1982 volcanic eruption of El Chichon, Geophys. Res. Lett., 10, 313-316.

Hofmann, D. J. and J. M. Rosen (1983), On the temporal variation of stratospheric aerosol size and mass during the first 18 months following the 1982 eruption of El Chichon, submitted to J. Geophys. Res.

Hofmann, D. J., J. M. Rosen, R. Reiter and H. Jaeger (1983), Lidar and balloon-borne particle counter comparisons following recent volcanic eruptions, J. Geophys. Res., 88, 3777-3782.

Iwasaka, Y., S. Hayashida and A. Ono (1983), Increased backscattered light from the stratospheric aerosol layer after Mt. El Chichon eruption; laser radar measurement at Nagoya (35°N, 137°E), Geophys. Res. Lett., 10, 233-235.

Iwasaka, Y. and K. Isono (1977), Lidar observations of the stratospheric aerosols at two different wavelengths, 0.6943 m, and 1.06 m, J. Atmos. Terr. Phys., 39, 117-120.

Kent, G. S., P. Sandland and R. W. H. Wright (1971), A second generation laser radar, J. Appl. Meteo., 10, 443-452.

Kerker, M. (1969), The Scattering of Light and Other Electromagnetic Radiation, Academic Press.

Klett, J. D. (1981), Stable analytical inversion solution for processing lidar returns, Appl. Opt., 20, 211-220.

Kneizys, F. X., E. P. Shettle, W. O. Gallery, J. G. Chetwynd, Jr., L. W. Abreu, J. E. A. Selby, R. W. Fenn and R. A. McClatchey (1980), Atmospheric transmittance/radiance: computer code LOWTRAN 5, Environmental Res. Paper 697, AFGL.

Letrere, J., P. Flamant and G. Megie (1983), Lidar survey of volcanic debris in the stratosphere following the 1982 eruption of El Chichon volcano, IAMAP Symp. R2, Hamburg.

Liou, K.-N. and R. M. Schotland (1971), Multiple backscattering and depolarization from water clouds for a pulsed lidar system, J. Atmos. Sci., 28, 772-784.

McCormick, M. P., J. D. Lawrence, Jr. and F. R. Crownfield, Jr. (1968), Mie total and differential backscattering cross sections at laser wavelengths for Junge aerosol models, Appl. Opt., 7, 2424-2425.

Reagan, J. S., J. O. Spinhirne, D. M. Byrne, D. W. Thomson, R. G. de Pena and Y. Mamane (1978), Atmospheric particulate properties from lidar and solar radiometer observation compared with simultaneous in situ aircraft measurements: a case study, J. Appl. Meteo, 16, 911-928.

Russell, P. B. and B. M. Morley (1982), Orbiting lidar simulations. 2. Density, temperature, aerosol, and cloud measurements by a wavelength-combining technique, Appl. Opt., 21, 1554-1563.

Russell, P. B., B. M. Morley, J. M. Livingston, G. W. Grams and E. M. Patterson (1982), Orbiting lidar simulations. 1. Aerosol and cloud measurements by an independent-wavelength technnique, Appl. Opt., 21, 1541-1553.

Spinhirne, J. D., J. A. Reagan and B. M. Herman (1980), Vertical distribution of aerosol extinction cross section and inference of aerosol imaginary index in the troposphere by lidar technique, J. Appl. Meteo., 19, 426-438.

Swissler, T. J., M. P. McCormick and J. D. Spinhirne (1983), El Chichon eruption cloud: Comparison of lidar and optical thickness measurements for October 1982, Geophys. Res. Lett., 10, 885-888.

Toon, O. B. and J. B. Pollack (1976), A global average model of atmospheric aerosols for radiative transfer calculations, J. Appl. Meteo., 15, 225-246.

Turco, R. P., R. C. Whitten and O. B. Toon (1982), Stratospheric aerosols: Observation and theory, Rev. Geophys. Space Phys., 20, 233-279.

Zuev, V. E. (1976), Laser light transmission through the atmosphere, in Laser Monitoring of the Atmosphere, E. D. Hinkley, editor, Springer-Verlag.

Zuev, V. E. and I. E. Naats (1982), Inverse Problems of Lidar Sensing of the Atmosphere, Springer-Verlag.

6. ACTIVE OPTICAL SOUNDING OF OZONE AND MINOR CONSTITUENTS IN THE
MIDDLE ATMOSPHERE: A REVIEW OF GROUND-BASED LIDAR MEASUREMENTS

G. Megie and J. Pelon

Service d'Aeronomie CNRS
BP 3-91370 Verrieres le Buisson
France

Continuous monitoring of ozone and other minor constituents in the earth's
stratosphere and troposphere is of particular interest in present day atmo-
spheric physics, as the potential modification of the ozone layer due to man's
activities can greatly modify the earth's environment and climate; The measure-
ment of th total ozone column content and vertical profile by means of the
ground based UV spectrometer network or by satellite-borne systems remains the
fundamental basis for global observations and trend analysis. However the re-
quired accuracy of these measurements will imply also in the future the opera-
tion of new systems which could provide a high degree of reliability in terms of
accuracy and absolute precision. Furthermore the very large variability of the
ozone number density in the troposphere and lower stratosphere and at the bound-
ary between these two regions and it interpretation in terms of horizontal and
vertical transport requires high spatial and temporal resolution measurements
which are presently beyond the possibilities of passive systems. The recent
development of powerful tunable laser sources has opened a new experimental
field for spectroscopic studies of atmospheric trace constituents. As part of
it, the differential absorption laser (DIAL) technique has now been used to pro-
vide vertical profiles of the ozone number density from the ground up to the 40-
50 km altitude level. The purpose of this paper is thus to give a detailed
analysis of the method itself and to describe the various systems presently in
operation. An analysis of their advantages and disadvantages with respect to
one another will also be presented and an overview of already obtained data will
be given. Perspectives in terms of ozone and other constituents monitoring
will be considered.

METHODOLOGY - THE DIAL TECHNIQUE

The basic principles of the differential absorption lidar technique have
been described by various authors (SCHOTLAND, 1964; BYER and GARBUNY, 1973).
Its application to ozone measurements for both UV and IR systems has been ana-
lyzed by MEGIE and MENZIES (1980). They concluded that as far as higher tropo-
spheric and stratospheric measurements are concerned the UV wavelength range is
the best candidate for ground based observations. Therefore only such systems
operating in the Hartley-Huggins bands of ozone between 280 and 320 nm will be
considered here. The forthcoming analysis will follow the study made by PELON
and MEGIE (1982a).

The usual lidar equation which relates the total number $N_{\lambda R}$ of backscat-
tered photoelectrons at wavelength λ from the cell at the range R and of thick-
ness ΔR to the number of emitted photons in the laser pulse at wavelength λ is
written as (Figure 1):

$$N_{\lambda R} = N_{e\lambda} \cdot \Delta R \cdot \frac{A}{R^2} \beta_{\lambda R} \cdot \eta \cdot \eta' \cdot \exp \left\{ - 2 \left(\tau_{\lambda R}^{o} + \tau_{\lambda R}^{e} \right) \right\} \qquad (1)$$

where $\beta_{\lambda R}$ is the atmospheric backscattering coefficient at wavelength λ and
range R; ΔR, thickness of the range cell corresponding to a time gate interval

of $\frac{2 \Delta R}{C}$ generally larger than the pulse duration Δt; A, receiver area; η, de-

tector efficiency; η', optical efficiency of the transmitter-receiver system;
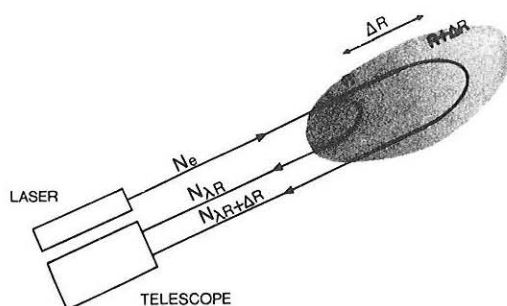
Figure 1. Principle of the DIAL technique.

$\tau^o_{\lambda R}$ integrated optical thickness due to the absorption by the constituent under study, i.e., ozone:

$$\tau^o_{\lambda R} = \int_0^R \sigma_{\lambda R} \, n_o \, (R) \, dR \tag{2}$$

where $\sigma_{\lambda R}$ is the ozone absorption cross section and $n_o(R)$ the ozone number density; $\tau^e_{\lambda R}$, integrated optical thickness excluding absorption by ozone; $\tau_{\lambda R}$, total integrated optical thickness $\tau_{\lambda R} = \tau^o_{\lambda R} + \tau^e_{\lambda R}$. If one considers the ratio of the two backscattered signals corresponding to two successive cells R_1, and $R_2 = R_1 + \Delta R$ for the same laser pulse one can write:

$$\Delta \tau^o_{\lambda 1} = \frac{1}{2} \ln \frac{N_{\lambda 1}}{N_{\lambda 2}} + \ln \frac{\beta_{\lambda 2} \, R_1^2}{\beta_{\lambda 1} R_2^2} - \Delta \tau^e_{\lambda 1} \tag{3}$$

where $\Delta \tau_{\lambda R}$ is defined as the local optical thickness of the constituent within the range cell R, $R + \Delta R$. Assuming the knowledge of the ozone absorption cross section $\sigma_{\lambda R}$ at this range, $\Delta \tau^o_{\lambda R}$ is then directly proportional to the average ozone number density within the same range cell as

$$\Delta \tau^o_{\lambda R} = \sigma_{\lambda R} \cdot n_o(R) \cdot \Delta R \tag{4}$$

Thus a one wavelength lidar measurement can lead to the determination of the ozone vertical profile from the laser backscattered signals $N_{\lambda 1}$, $N_{\lambda 2}$ if the last two terms of equation (3) are known. If one assumes a pure molecular scattering atmosphere (no aerosol particles) $\beta_{\lambda R}$ and $\Delta \tau^e_{\lambda R}$ are then proportional to the atmospheric number denisty as the Rayleigh scattering process is the only one contributing to the laser light scattering. Using thus an atmospheric model or the atmospheric parameters as given by local radiosonde measurements such a technique can be used for ozone profiling (UCHINO et al., 1979). However aerosol particles are always present in the troposphere and in the lower stratosphere around 20 km and their abundance can be greatly increased by more than one order of magnitude during the one or two years following large volcanic eruptions (MCCORMICK et al., 1983). Single wavelength measurements can then not be considered as fully reliable for any atmospheric situation and one has to use a second wavelength λ_2 different from λ_1 to discriminate the ozone absorption from other potential interfering species. Writing equation (3) for a second wavelength λ_2 and taking the difference in between leads then to:

$$\Delta \tau_{1R} = \Delta \tau_{2R} = \ln \frac{N_{22} N_{11}}{N_{12} N_{21}} + \ln \frac{\beta_{12} \, \beta_{21}}{\beta_{11} \beta_{22}} \tag{5}$$

where the first subscript refers to the wavelength and the second one, when present, to the range.

The signal-to-noise ratio $S_{\lambda R}$ for a single wavelength measurement is given, in the case of an incoherent detection scheme and assuming that the fluctuations of the various sources of the photodetector current are governed by Poisson statistics by

$$S_{\lambda R} = \left(\frac{N_{\lambda R}}{(N_{\lambda R} + N_B + N_D)} \right)^{1/2} \qquad (6)$$

where N_B is the number of background photons and N_D the square of the photo-detector dark current fluctuations within the time gate interval $2\Delta R/c$.

In the usual derivation of the DIAL method, one considers that the wavelength variations of $\beta_{\lambda R}$ and $\tau^e_{\lambda R}$ between λ_1 and λ_2 contribute only to the systematic error term ε_2 (see below)[1] (see, for example, SCHOTLAND, 1974) so that (3) can be rewritten as

$$\Delta\tau^o_{1R} - \Delta\tau^o_{2R} = 1/2 \, \ell n \, (\frac{N_{22}N_{11}}{N_{12}N_{21}}) \qquad (7)$$

The average ozone number density $n_o(R)$ within the range cell R, $R + \Delta R$ is given by

$$n_o(R) = \frac{\Delta\tau^o_{1R} - \Delta\tau^o_{2R}}{(\sigma_1 - \sigma_2)\Delta R} \qquad (8)$$

The relative uncertainty $\varepsilon = \frac{\delta n_o}{n_o}$ depends on a large number of interrelated parameters and can be expressed as the sum of two terms, $\varepsilon_1 + \varepsilon_2$.

(1) ε_1 is the statistical error due to the signal and background noise fluctuations and is related to the signal uncertainty by

$$\varepsilon_1 = \frac{\kappa}{2\tau^o_R \sqrt{N_o}} \quad (f_1^2 + f_2^2)^{1/2} \qquad (9)$$

where

$$K = \frac{\tau^o_R}{\Delta\tau^o_R} = \frac{\int_o^R n_o(R)dR}{n_o(R) \, \Delta R} \qquad (10)$$

is wavelength independent as far as one considers that the ozone absorption cross sections are altitude independent; One can also notice that the product $K\Delta R$ does not depend on the range resolution of the lidar system (1) τ^o_R is the differential integrated optical thickness: $\tau^o_R = \tau_{1R} - \tau_{2R}$; (2) $N_o = A\Delta R n n'/R^2$ is the wavelength independent part of $N_{\lambda R}$; (3) f_λ is a wavelength dependent function given by

$$f_\lambda^2 = \frac{e^{2\tau_{\lambda 1}}}{\beta_\lambda N_{e\lambda} P_\lambda e^{-2\tau_{\lambda 1}}} \{(1 + e^{2(\tau_{\lambda 1} + \tau^o_\lambda/K)})$$

$$+ X^o_\lambda (1 + 4e^{\mu_{\lambda 1}} e^{4(\tau_{\lambda 1} + \tau^o_\lambda/K)}\} \qquad (11)$$

$$\mu_\lambda = \ell n \{(1 + \frac{\Delta R}{R}) (\frac{\beta_{\lambda 1}}{\beta_{\lambda 2}})^{1/2}\} + \Delta\tau_\lambda^e \qquad (12)$$

$$X^0_\lambda = \frac{N_B + N_D}{N_{\lambda 1}} \, e^{-2} \tau_\lambda \, e \qquad (13)$$

is independent of the ozone absorption; and (4) P_λ is the number of laser pulses emitted at the wavelength λ.

(2) ε_2 is a systematic error due to the wavelength dependence of the scattering and absorbing (other than ozone) properties of the atmospheric medium which has been neglected in the derivation of equation (5). A general expression of ε_2 has been given by MEGIE and MENZIES (1980) as

$$\varepsilon_2 = \frac{K \Delta R}{\tau_R^0} \cdot G(R, \lambda) \cdot \frac{\Delta \lambda}{\lambda_1} \qquad (14)$$

$G(R, \lambda)$ is a function of range and wavelength depending on the scattering and extinction properties of the atmospheric gas and particles, written as

$$G(R, \lambda) = 2 \left(1 - \frac{m}{4}\right) \frac{1 - r_{\lambda R}}{(r_{\lambda R})^2} \left(\frac{1}{h} - \frac{1}{H}\right) + m\alpha_\rho + {}^4\alpha_M \qquad (15)$$

Using these expressions of the various uncertainties an error analysis of the DIAL measurement taking into account both experimental parameters and atmospheric characteristics can now be performed.

A minimum value of ε can be obtained for a given range by an appropriate choice of the laser emission wavelength λ_1 and λ_2. In the general case, *this optimization of the DIAL* measurement requires a numerical computation to include the experimental parameters of the lidar system and the atmospheric parameters related to molecular and aerosols scattering and ozone absorption. The description of such computations are beyond the scope of this paper; however, a simple analysis can be made by considering separately the lower (0-2 km) and higher altitude ranges, as the relative contribution of the molecular and aerosol extinctions will be different in these two cases. Whereas this study is mainly devoted to the middle atmosphere, one has also to consider the lower tropospheric levels as far as a determination of the ozone total content is to be undertaken.

(a) Boundary Layer and Lower Tropospheric Measurements

Below 2 km, ε_2 becomes the most important error term as the molecular and aerosol extinctions are large and as the backscattered signal is intense enough so that the statistical error ε_1 can be reduced below 10^{-2}. By using the extinction values given by ELTERMAN (1968), the differential molecular extinction for a wavelength interval $\Delta\lambda = \lambda_1 - \lambda_2 = 5$ nm can be as large as 0.1 in the wavelength range considered (270 - 310 nm) and can thus not be neglected. A correction has to be made by measuring the ground-level pressure and temperature to derive the atmospheric density in the first kilometers. An uncertainty of a few percent in this determination will reduce the error due to molecular extinction to less than 5×10^{-3}.

The remaining error in ε_2 is then related to the aerosol extinction, and a further distinction should be made between ozone measurements performed in rural (nonpolluted) or urban areas.

(1) Rural Areas: We have calculated the variation of ε_2 as a function of wavelength and range by using aerosol particle concentrations as given by ELTERMAN (1970) and typical of rural conditions. These variations are given in Figure 2 for different values of λ_1 and $\Delta\lambda = \lambda_2 - \lambda_1$ for a vertical range R = 1 km. The optimization of λ_1 and $\Delta\lambda$ in the wavelength range 265-285 nm leads to values of ε_2 less than 2%, which can be further reduced if needed by using the experimental procedure adopted for urban or polluted areas and described in the following subsection.
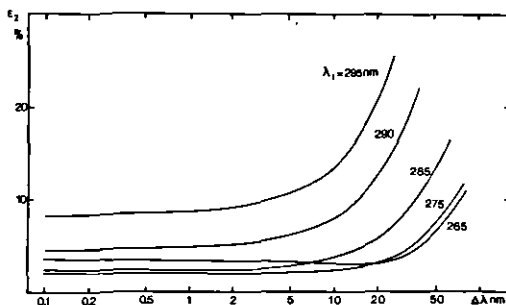
Figure 2. Variation of the systematic error ε_2 as a function of the wavelength difference $\Delta\lambda = \lambda_2 - \lambda_1$ for different values of the absorbed wavelength λ_1 for a range R = 1 km (rural aerosol model).

(2) Urban or Polluted Areas: Owing to the large aerosol particle concentrations in the boundary layer, ε_2 can reach values as large as 10^{-1}, and a complementary measurement will be needed to eliminate the aerosol differential extinction. This can be done, for example, by using three wavelengths in the 265-285 nm range with the same wavelength interval $\Delta\lambda = 5$ nm. The aerosol differential extinction can then be subtracted, if one assumes a linear variation for the aerosol scattering and extinction properties over 10 nm. This will avoid the use of a theoretical model that depends on the nature and shape of the particles. Taking into account the experimental errors and the modeling uncertainties, the estimated upper limit of ε_2 will then be reduced to 2-3%.

(b) Upper Tropospheric and Stratospheric Measurements

Above 2 km, ε_2 decreases rapidly with altitude, as represented in Figure 3, for average values of higher tropospheric and stratospheric aerosol content. The choice of the operating wavelengths is then determined by calculating the minimum value of ε_1 as a function of the three interrelated and wavelength-dependent parameters τ_{1R}^0, τ_{2R}^0, $\Delta\lambda$. For a given value of the larger absorption wavelength λ_1, the minimum is obtained for $\tau_2^0 = 0$. This condition cannot be experimentally achieved in the case of the ozone UV absorption bands, which present continuous absorption features over a wide wavelength range. Thus the optimization procedure will consist of first calculating the value of τ_{1R}^0 and thus λ_1, corresponding to the minimum value of ε_1 for $\Delta\lambda = \infty$ (i.e., $\tau_{2R}^0 = 0$) and then evaluating the decrease in accuracy that results from the choice of a finite value of $\Delta\lambda$ compatible with both the experimental constraints and a maximum value of ε_2 as calculated above. MEGIE and MENZIES (1980) have shown that the optimum value of τ_{1R}^0 is 1.28, in the case of a shot-noise limited signal when one neglects the off line (λ_2) absorption (for ozone measurements this corresponds to large values of $\Delta\lambda$). We have represented on Figure 4 the variations of ε_1 relative to its minimum value as a function of $\Delta\lambda$. The values of the atmospheric density as a function of altitude required for the calculations were taken from U.S. Standard Atmosphere (1976). By comparing these variations with the ones obtained for ε_2 (Figure 2), one can see that the optimum wavelength interval has to be in the range 5-15 nm. The final choice of 5 nm has been adopted by considering the maximum tuning range of the laser systems and the possibility of using several wavelength pairs for the measurement within this range.

The optimum value of λ_1 as determined above is only valid for a given altitude range R. From an experimental point of view, it seems impossible to
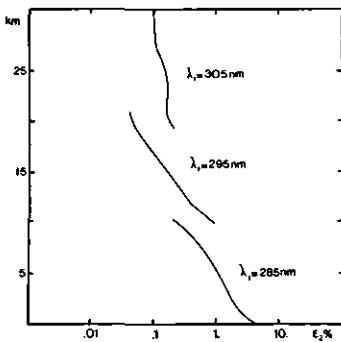
Figure 3. Variation of the systematic error ε_2 as a function of altitude for $\Delta\lambda = 5$ nm and the three values of λ, that optimize the statistical error ε_1.
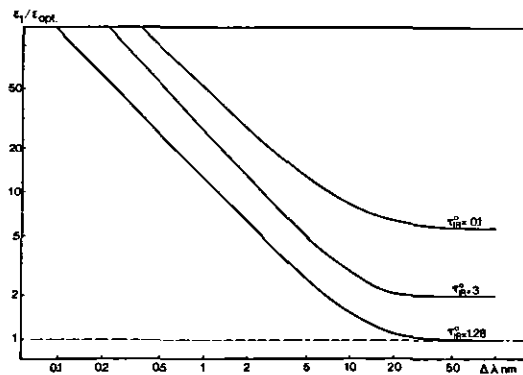

Figure 4. Variation of the statistical error ε_1 as a function of $\Delta\lambda$ for different optical thicknesses τ_{1R}^o corresponding to different wavelengths λ, or ranges. ε_1 is normalized to its minimum value.

use as many wavelength pairs as the number of altitude levels of the measurement. Therefore, we have calculated the altitude variations of ε_1 (λ_1, $\Delta\lambda$) by using the mean ozone profile of KRUEGER and MINZER (1976) typical of middle latitude regions. These relative variations represented on an arbitrary scale, corresponding to experimental conditions, are given on Figure 5 for three wavelength pairs (λ_1, λ_2). The shorter wavelengths are used to probe the lower altitude levels, and the useful range of a given pair can be extended up to 7-8 km so that the measurement accuracy stays between its optimum value ε and 1.2ε.

(c) Upper Stratospheric Measurements and Ozone Total Content

Attempts to measure ozone number density above its maximum using a ground-based UV system are complicated by the high extinction of the laser emitted light due to the absorption in the lower altitude levels. Furthermore, the rapid decrease in the ozone number density above 28 km requires very rapidly increasing acquisition times for the measurement. The choice of higher wavelengths (wavelength pair B width $\lambda_1 = 305.8$ nm and $\lambda_2 = 310.8$ nm) allows one to overcome these major problems and to probe the upper levels with still appropriate integration times for geophysical applications (PELON and MEGIE, 1982b).
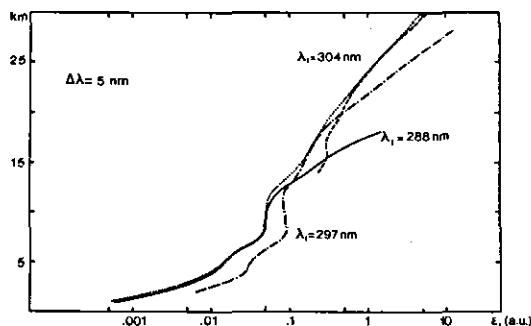
Figure 5. Variation of the statistical error ε_1 as a function of height for three different range optimized wavelengths. The dotted curve represents the optimum value of ε_1 for each altitude.

The ozone total content can also be derived from the measured vertical distribution using a combination of the two wavelength pairs (pair A at lower altitudes - 288-294 nm; pair B at higher levels). Various sources of uncertainty have to be considered:

(1) The statistical fluctuations of the detected signals which constitutes the major error source estimated to be $\leq$ 10 Dobson units in a 95% confidence interval. This error can be further reduced by increasing the acquisition time.

(2) The Rayleigh extinction by the atmospheric gas molecules: this uncertainty is reduced below 0.1 Dobson unit by correcting the lidar data using the pressure, temperature and humidity distribution measured by radiosonde.

(3) The Mie differential extinction by aerosol particles: this uncertainty which can be large at lower altitudes is reduced when using the A wavelength pair to measure the total content up to 15 km which leads to high ozone optical thickness as compared with the aerosol contribution. It can be estimated to be $\leq$ 1 Dobson unit whereas the use of the B pair in the same altitude range will have resulted in a one order of magnitude higher uncertainty.

(4) A systematic bias due to the discrepancies between the two wavelength pairs and the absolute cross-sections determinations. A comparison has been made as both wavelengths pairs A and B can be used to measure the ozone number density between 13 and 18 km: considering several profiles obtained during successive nights a systematic bias can be detected showing ozone values 2% higher when measured using the A pair. This discrepancy can be related (1) to uncertainties in the absolute values of the cross sections and (2) to the very high spectral resolution of the laser measurements as compared with the spectrometric ones. Due to experimental uncertainties, the detected bias between the A and B pairs could only be considered as estimations in this first approach.

(5) A constant value of 10.4 Dobson units which corresponds to the average ozone content above 40 km (KRUEGER and MINZER, 1976); the 2 σ-standard deviation results in a 3.2 Dobson units uncertainty.

(d) Interference With Other Absorbing Gases

Two minor constituents have to be considered as potential interference absorbers in the 300 nm wavelength range;

(1) The sulfur dioxide: SO_2 absorption spectrum presents a band structure with a spectral width for the individual features of 1 nm (THOMPSON et al., 1975). The differential cross section between an absorption maximum and a minimum ($\Delta\lambda \sim 1.5$ nm) can reach 10^{-18} cm^2. In the case of large amounts of SO_2 in the troposphere over urban areas, the corresponding differential absorption coefficient can reach values as large as 0.1 km^{-1} at the ground level (STEWART et al., 1978). To avoid any interference, the two laser wavelengths λ_1 and λ_2 should be chosen so that the SO_2 differential absorption vanishes. This choice is always possible within 0.3 nm of a predetermined wavelength.

(2) The nitrogen dioxide: NO_2 presents in the same wavelength range absorption bands with cross-section values of the order of 10^{-19} cm^2 (BASS et al., 1976). The differential cross section for a 5 nm wavelength interval is then $2.5\ 10^{-20}$ cm^2. Its results, if one considers the higher mixing ratios typical of urban or polluted areas (~ 0.3 ppm (SEINFELD, 1975)), in a differential absorption coefficient of 0.015 km^{-1} at the ground level. The corresponding error on the ozone measurement will be 2% and can be taken care of by the same procedure as for SO_2.

    In any case, the high altitude measurements will not be affected due to the very low atmospheric content in these interfering species above the boundary layer.

(e) Temporal Variations of the Scattering Medium

    If the two wavelengths are not simultaneously emitted, variations in the optical properties of the scattering medium might occur between the two laser shots, owing to atmospheric transport. This will be of particular importance at lower altitudes if the aerosol content is high or at the tropopause level in the presence of cirrus clouds. Thus, even if the integration time required for the measurement is large, the switching between the two wavelengths will have to be made with a time constant much smaller than the constant characteristic of the dynamical transport. The experimental system will have to take this requirement into account (see section on "Results"), and the possible design of a dual cavity laser emitting simultaneously at two wavelengths should also be investigated.

(f) Altitude Dependence of the Ozone - Absorption Cross Sections

    The value of the absorption cross sections that we used to derive the ozone number densities are taken from INN and TANAKA (1953). The spectral resolution of their measurements is lower than the resolution of the lidar measurements as given by the laser linewidth ($\sim 50,000$). The assumption has thus to be made that the mean value of the absorption cross section is the same over these different wavelength intervals. This is not experimentally confirmed, but the precise knowledge of the emitted wavelength (see the following section) will also allow one to correct data from this systematic error if high resolution absorption spectra of ozone become available in this wavelength range.

    The ozone absorption cross sections are temperature dependent mainly in the Huggins bands above 310 nm where variations as large as a factor of 2 can be observed for a 100 K temperature difference. Low resolution measurements (VIGROUX, 1953) have shown that the absorption minima are more sensitive than the maxima to the temperature variations. In the Hartley bands, the relative temperature variation of the cross section is much lower: 1% for a 10 K variation (VIGROUX, 1953). Therefore, the use of an atmospheric model to take into account this temperature dependence of the absorption cross section and an a

posteriori control using the radiosonde measurements of the nearby meteorological stations will reduce this uncertainty on the ozone profile to less than 0.5%.

EXPERIMENTAL SYSTEMS

The system evaluation conducted in the previous section shows that the measurement of the ozone number density profile by the DIAL technique is achievable using presently available laser systems with a temporal resolution compatible with scientific geophysical objectives. Two altitude domains might be considered

– below 30 km and due to the presence of aerosol layers in the troposphere and stratosphere a dual wavelength operation is required with a wavelength interval between the two emitted laser lines not larger than 5 nm

– above 30 km where the atmosphere can be generally considered as purely molecular – whereas after large volcanic eruptions aerosol layers have been observed up to 40 km – a single wavelength lidar might give information on the ozone profile. However such a determination will imply the use of an atmospheric density model which can lead to uncertainties in the ozone values especially during perturbed conditions. This difficulty might be overcome using a second laser emitted wavelength with the possibility of an increased wavelength interval.

(a) Transmitter

The design of an experimental system must also take into account the available laser sources; two types of laser are presently potential candidates for such a system:

– frequency-doubled dye lasers pumped by a Nd^{3+}Yag laser which present the great advantage to be tunable over a broad wavelength range but are somewhat limited in terms of available output power,

– excimer lasers which are frequency fixed on a given transition of the active medium but can emit very large average and peak powers.

Both lasers have been used for the measurements of atmospheric ozone. The first and presently only ground-based lidar system using dye lasers has been operational at the Observatoire de Haute Provence since 1977 (MEGIE et al., 1977, PELON and MEGIE, 1982a, PELON and MEGIE 1982b) (Figure 6). The active part of the transmitter is a laser pumped frequency-doubled dye laser. The pump laser is a Nd^{3+}Yag laser (Quantel model 480) emitting an energy of 750 mJ at 1.06 µm with a repetition rate of 10 Hz. This IR emission is frequency doubled with an efficiency of 40% resulting in an available pump energy of 300 mJ at 0.53 µm. The transversly pumped dye laser (Jobin Yvon model HPHR) includes one oscillator cavity and three amplifier stages (BOS, 1981). To cover the wavelength range from 570 to 620 nm, two dye solutions are used as the active medium (1) 570–600 nm a 5×10^{-4}M/1 solution of Rh G in water + 5% ammonix; (2) 590–620 nm a 5×10^{-4}M/e solution of Rh 610 in water and 5% ammonix.

The energy conversion efficiency is 40% corresponding to an output energy of 120 mJ. The wavelength selection and spectral narrowing of the emitted laser line are made by using a 2750 grooves mm^{-1} grating at Littrow incidence. The emission characteristics of this laser are summarized in the Table 1. To adapt these characteristics to the requirements brought out from section 2, two experimental achievements remain to be made.

(1) The output wavelength has to be converted to a value in the near UV wave-
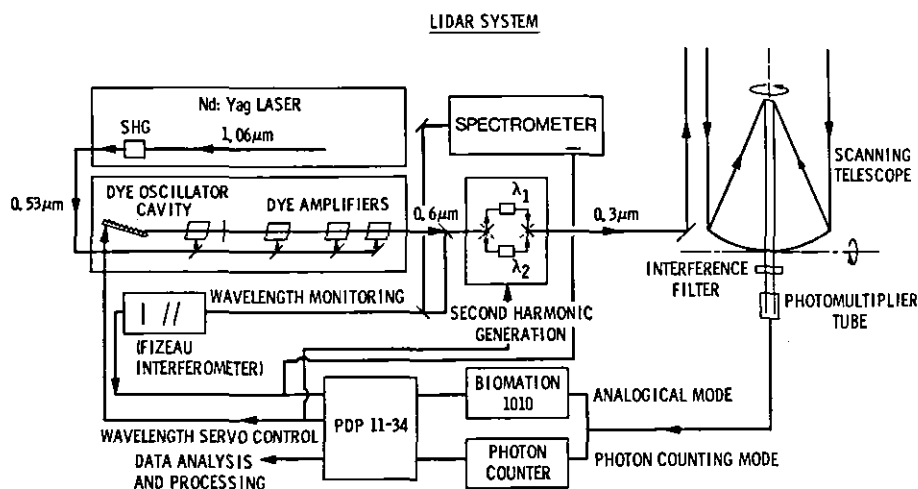
Figure 6. Synoptic diagram of the lidar system in
operation at the Observatoire de Haute Provence.

TABLE 1

CHARACTERISTIC OF PRESENTLY OPERATIONAL LIDAR SYSTEMS

Observatoire de Haute Provence 44 N, 5E, (PELON et al. 1982a)

Emitter (Dye laser, dual wavelength)		Receiver	
Energy (280 - 300 nm)	40 mJ	Telescope diameter	36 or 80 cm
Pulse duration	12 ns	Receiver bandwith	70 nm or 3 nm
Repetition rate	10 Hz	Field of view	1 mrd
Emission linewidth	5 pm	Vertical resolution (maximum)	150 m
Beam divergence	0.5 mrad.	Altitude range	0-40 km

Fukuoka 33N, 130E (UCHINO et al. 1983)

Emitter (XeCl, monowavelength)		Receiver	
Energy (308 nm)	50 mJ	Telescope diameter	50 cm
Pulse duration	16 ns	Receiver bandwith	20 nm
Pulse repetition	5 Hz	Field of view	750 m
Beam divergence	1 mrad	Altitude range	15-25 km

Zugspitze (ROTHER et al. 1982)

Emitter (XeCl, dual wavelength)		Receiver	
Energy (308 nm)	130 mJ	Telescope diameter	60 cm
Pulse duration	15ns	Receiver bandwith	1 nm
Pulse repetition rate	40 Hz		
Emission Linewidth	0.7 nm	Vertical resolution (maximum)	15 m
		Altitude range	30-45 km

length range by second harmonic generation using a KDP crystal. Owing to the high peak power of the fundamental emission (20 MW) the energy conversion efficiency is close to 35%. The output energy between 285 and 310 nm is equal to 40 mJ.

(2) The laser system has to emit sequentially two wavelengths. It is designed so that the full sequence of wavelength switching is automatic. The output wavelength of the dye laser fundamental emission is monitored by using a spectrometer and a Fizeau interferometer, giving patterns recorded on diode arrays. A computer-controlled servo mechanism (CAHEN et al., 1981b) is used to both ensure the stability of the output wavelength from λ_1 to a second pre-programmed value λ_2. To allow this switching to be fast ($t \leq 1$ s) and to avoid the experimental difficulties connected to the accuracy of the doubling crystals positioning as a function of wavelength, two KDP crystals are used that are preset to the optimum value of the phase matching angle for λ_1 and λ_2. The laser beam is then mechanically switched from one crystal to the other, depending on the incident wavlength. When a second wavelengths pair has to be used, the values of λ_1 and λ_2 as programed for the servocontrol loop are changed so as the crystal angles. By using this system, the stability of the laser emission wavelength is better than 1 pm and the switching operation takes place in less than 0.5 s. As the laser divergence is smaller than 5 x 10^{-4} rad, no transmitting optics is used and the beam is sent directly upward by using a total reflection prism.

Excimer lasers for ozone measurements have been first used by UCHINO et al., in 1980. They use a Xe-Cl laser emitting at 308 nm. The maximum output energy was 128 mJ/pulse for the first system developed with a rather low repetition rate of 0.1 pulse/s so that the average power is of the same order of magnitude as for the dye laser. However improvement of these systems has lead to a XeCl laser with an output energy of 130 mJ and a repetition rate up to 100 Hz (ROTHE et al., 1982). The Japanese group has only performed up to now single wavelength measurements. Dual wavelength systems are presently developed which use either:

- stimulated Raman scattering generation. The 308 nm radiation is focussed into a high-pressure methane cell to generate the reference line (off absorption). The wavelength interval is then equal to 2880 cm^{-1} leading to an emission at 338 nm which is absorbed less significantly by ozone. The conversion efficiency is 15% for a pressure of 35 atm and a focussing length of 125 cm. The two lines can be emitted simultaneously.

- the third harmonic of a Nd^{3+}Yag laser at 355 nm which requires an additional laser source. The large wavelength interval (47 nm) allows only measurement above 25 km if aerosol free conditions can be assumed.

As a summary the characteristics of the lidar systems in operation are given in Table 1.

Optical Receiver Electronic Processing and Data Acquisition

These parts of the lidar system are similar for the various laser system which have been used up to now (see Table 1). The following description refers to the system presently in use at the Observatoire de Haute Provence.

The backscattered signal is collected by either a 36 cm or a 80 cm diameter in a Cassegrain configuration. The distance between the emitting point and the telescope axis can be varied, depending on the altitude range of the observations to increase or decrease the range at which the fields of view of the transmitter and the receiver begin to overlap. This will avoid saturation of the multiplier tube from the low altitude backscattered signals. The telescope

field of view is adjusted by using a remotely controlled iris and can thus be reduced to its limit value compatible with the laser divergence. The spectral bandwidth of the receiver can be reduced by either using a wide-band inter- ference filter (70 nm) or several narrow bandwidths interference filters (3 nm) which can be automatically changed when the laser emission wavelengths are changed. The signal is detected by using a photomultiplier tube. The dynamic range of the backscattered signal between the ground and 30 km can be as large as 10^6 according to equation (1). Two acquisition modes are thus used. (1) For the lower altitude range the electrical signal delivered by the PMT is analysed by using a transient waveform recorder (Biomation 1010) with a sampling frequency of 10 MHz, which corresponds to a maximum altitude resolution of 15 m. A 10-bit converter is used for the analog-to-digital conversion. (2) For the altitude levels above 10-12 km, a 256 channels photon counter is used in paral- lel with the transient recorder, so that the two acquisition modes overlap with respect to the altitude range. The timegate of the photon counter can be varied from 1 to 8 μs and the maximum altitude resolution is thus 150 m.

The data provided by the two acquisition systems are then sent to a PDP 11-34 computer and are stored on a floppy disk. A presummation of the single laser shot signals is made to reduce the volume of stored data. The PDP 11-34 computer is used to control the full sequence of a DIAL measurement. The ex- perimental parameters (acquisition time, altitude resolution, values of the various laser wavelengths, switching time, etc.) are typed in to start auto- matically the sequence of laser firings. During the experiment, the laser energy and emission wavelength are continuously monitored, and the data acquisi- tion takes place only if all these parameters are within the range of predeter- mined values.

RESULTS

The reliability of the various systems presented here for ozone monitor- ing has been tested on an operational basis for the last three to four years. As a result the DIAL technique has proved its ability to provide altitude pro- files of ozone number density from the ground up to 40 km at the present.

The measurements performed at the Observatoire de Haute Provence using dye lasers as the transmitter are obtained in successive steps within integra- tion times of the order of 1 hour. Figures 7 and 8 are profiles corresponding to the altitude ranges 0-17 km and 15 km-40 km and are obtained sequentially. This experimental procedure results from the wavelength optimization as a function of height. The first profile is obtained within 15 min with a rela- tive accuracy better than 5% for a vertical resolution of 450 m. The higher altitude profile requires an integration time of about 45 min. The vertical resolution decreases from 1 km at 25 km to 3 km at 35 km and above. Here again the 1 σ standard deviation is better than 5% up to 25 km and decreases down to 20% at the uppermost level. Several comparisons with in situ measurements have been performed especially during the June 1981 Intercomparison campaign held at Gap (M.L. CHANIN, 1983). The results show a good agreement between in situ and remote sensing instruments (MEGIE and PELON, 1983, PELON and MEGIE, 1983).

Similar comparisons with balloon-borne ozonosondes have also been per- formed by UCHINO et al. (1980) covering the altitude range of XeCl measurements between 15 and 30 km. Here again the agreement is rather good (Figure 9). The XeCl laser developed by ROTHE et al. (1982) has been implemented at a high altitude (3 km asl) station at the top of the Zugspitze in the German Alps. This system is in operation since the winter of 1982 and will provide high alti- tude measurements between 30 and 45 km.

One of the first applications of the ground-based lidar system for ozone *measurements is the monitoring of the ozone vertical distribution on a routine*
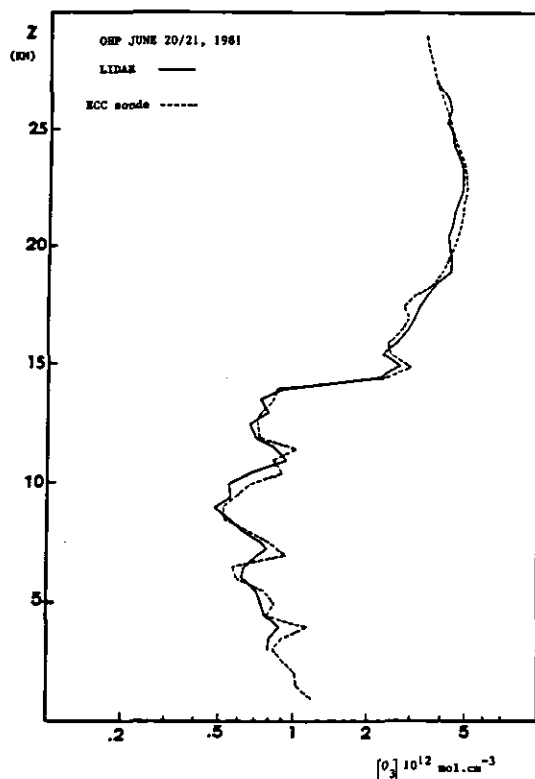
Figure 7. Vertical ozone distribution as measured
by an ECC sonde and lidar at the Observatoire
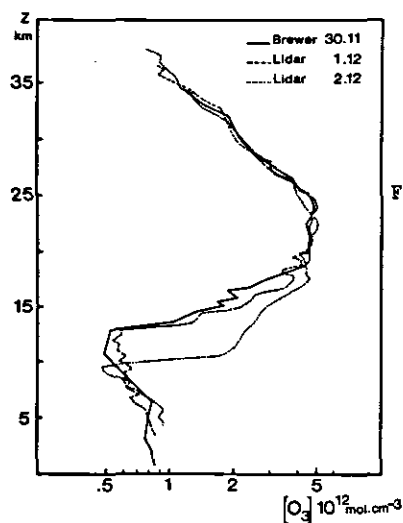de Haute Provence (June 1981).

Figure 8. Ozone concentration profile measured by
lidar at the Observatoire de Haute Provence
(44°N, 50E) on 1 December 1981 (dashed line) and
2 (dotted line) at 0 h UT compared with the profile
obtained by a Brewer Mast Sonde launched at
Biscarosse (44°N, 0,5°W) on 30 November 1981 at
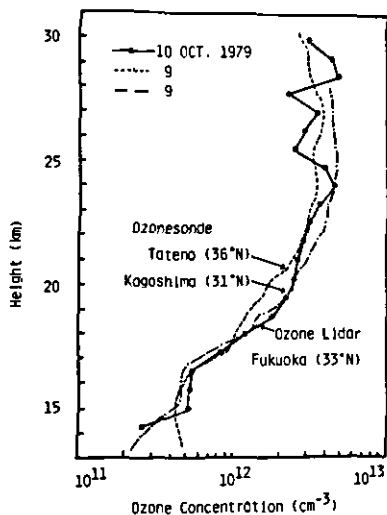13 h 30 UT (Solid line).

Figure 9. Comparison between Xe-Cl lidar and ozonesondes (from UCHINO et al., 1980).

basis. The altitude extension of the lidar measurements up to the 45 km level will allow the study of long term ozone variations in a region where the potential effects of manmade activities are maximum due to catalytic cycles involving more particularly chlorine species. With respect to passive systems presently in use the lidar system presents for such observations several advantages:

- the control of the emitting source which results in a possible independent autocalibration of each system,

- the possibility by a careful choice of the emitting wavelength to avoid interferences by aerosols, sulfur dioxide or nitrogen dioxide

- the direct altitude resolution of the system.

Several studies have also been undertaken using lidar data which refer to various temporal and spatial scales. For example short term variability of the ozone number density below 16 km has been observed by PELON et al. (1982a) as represented in Figure 10 where successive profiles recorded during the night show maximum abundance at 10 km occurs betwen 9 PM and 2:30 AM on the following day. Such ozone variations, already observed by other techniques (DUTSCH, 1979) are related to horizontal transport at the tropopause level. For such studies the temporal continuity in the observations which characterizes the lidar systems is of great advantage as it allows a determination of the horizontal extension of such ozone structures with a very high spatial accuracy.

Large variations in ozone number density are also observed on a day-to-day basis. They are related to larger scale horizontal transport. As an example, the vertical ozone profile recorded during the night of July 9, 1980 (Figure 11) shows the presence of a very large ozone bulge with concentration up to 3×10^{12} cm^{-3} at the 10-km altitude level. For comparison, the average profiles observed during the beginning of July 1980 are plotted on the same figure. Such increases of the ozone concentration at the tropopause level have been observed on several occasions during field experiments. The altitude of the ozone bulge may vary from 10 to 12 km down to 6 to 8 km as on March 11, 1981; in this latter case the peak concentration decreases when the bulge is observed at lower altitudes. From the meteorological network data, one can show that
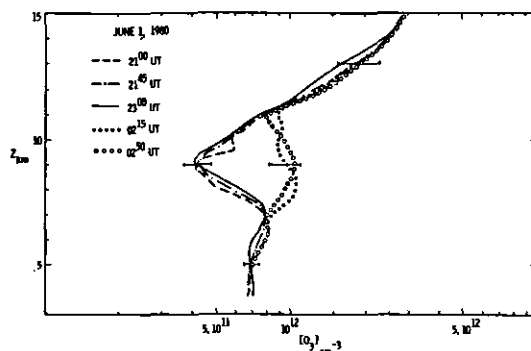
Figure 10. Short term variability of ozone
concentration profiles as recorded during
the night of June 1-2, 1980 at the Obser-
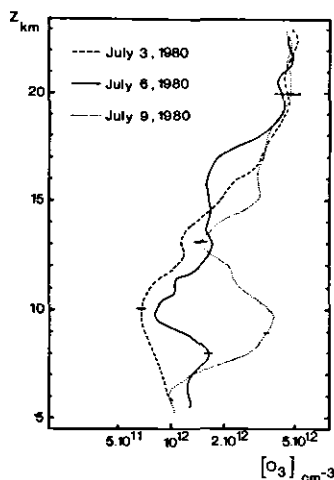vatoire de Haute Provence.

Figure 11. Ozone concentration profiles
recorded during July 1980 and showing
evidence for an ozone bulge occurrence
on July 9.

these increases are systematically correlated with a 2-3 km decrease of the
tropopause height corresponding thus to the presence of warmer air at these
levels over southern France. A detailed analysis of one of these situations
(July 1980) has already been performed by using the meteorological charts at
various altitude levels in the troposphere and lower stratosphere (PELON et al.,
1981). For such events also the potentiality of a ground-based system can be
fully utilized as shown by the study of UCHINO et al. (1983) which gives evi-
dence for a high positive correlation between the ozone column density in the
altitude range of 15-25 km with the total ozone observed by a nearby Dobson
spectrophotometer (Figure 12). Other correlations have also been studied by the
same group such as a high positive correlation with temperature at 17.25 km;
all these features being closely related to dynamical movements at the tropo-
pause height (PELON et al., 1983).

The emphasis in this presentation has been put on ozone measurements.
When considering minor species in the stratosphere the relative abundance
ratios vary from the ppm range down to the ppb or even ppt ranges. As shown
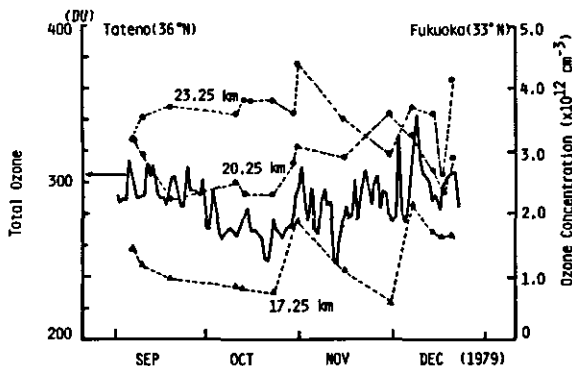
Figure 12. Comparison between the total ozone
observed by the Dobson spectrophotometer at
Tateno and the ozone mean concentrations
heights measured by XeCl Lidar (UCHINO et al.,
1980).

above the ability of a ground-based active system measurement is strongly de-
pendent on the local absorption to be measured. If one considers that the ab-
sorption cross sections are generally in the 10^{-18} cm^2 range an abundance
ratio of 1 ppm at 30 km will lead to a 1-km optical thickness of 2.5%. A
reasonable detection limit for lidar absorption measurement can presently be
fixed around 2.5% for a two-way absorption which limits the potential measure-
ments to minor constituents in the 50 to 100 ppb range. Furthermore, the ver-
tical profile of the constituent under study is also to be taken into account
as the detection at a given altitude level is strongly dependent on the absorp-
tion below. For example water vapor measurements in the stratosphere from the
ground which could be possible in the 940 nm spectral region, are prohibited by
the very large tropospheric content. Besides ozone the only stratospheric con-
stituent which has been measured up to now is nitrogen dioxide using a laser
source which emits at 440 nm (BUCCHIA and MEGIE, 1983). However only the NO_2
total content between 20 and 40 km has been directly measured, the altitude
profile being derived only using inversion techniques. The potentiality of
active systems for trace species measurement in the stratosphere is however much
larger if one considers airborne, balloon-borne or space-borne systems.

CONCLUSION

Due to the rapid development of powerful laser sources in the UV wave-
length range which includes Nd:Yag pumped dye lasers and excimer lasers, ground-
based lidar systems are now operational for ozone monitoring in the troposphere
and the stratosphere. They can provide a direct measurement of the ozone ver-
tical distribution from the ground up to the 45 km altitude level. Such
systems will certainly play an important role in the determination of long term
ozone trends in the photochemical region and thus in the evaluation of anthropo-
genic effects in the earth's environment. Furthermore, the unique capacities
of active lidar systems in terms of high temporal and spatial resolution and
measurements continuity, allow the observation of the ozone variations at vari-
ous time and space scales which are of importance in such present day areas of
interest as troposphere-stratosphere exchanges, long-range transport and global
budget of ozone or correlations between ozone number densities and other atmo-
spheric parameters. Considering the already operational character of the lidar
systems, they will constitute in the very near future the basis for the develop-
ment of a new ground-based network for middle atmospheric observations of ozone
and other trace gases.

REFERENCES

Bass, A. M., A. E. Ledford, and A. H. Laufer, Extinction coefficients of NO_2 and N_2O_4, J. Res. NBS, 80A, 143, 1976.
Bos, F., Versatile high-power single-longitudinal-mode pulsed dye laser, Appl. Opt., 20, 1886, 1981.
Bucchia, M., and G. Megie, Ground-based active remote sensing of the nighttime NO_2, Annal. Geophys., in press, 1983.
Byer, R. L. and M. Garbuny, Pollution detection by absorption using Mie scattering and topographic targets as retroreflectors, Appl. Opt., 12, 1496, 1973.
Cahen, C., J. P. Jegou, J. Pelon, P. Gildwarg, and J. Porteneuve, Wavelength stabilization and control of the emission of pulsed dye lasers by means of a multibeam Fizeau interferometer, Rev. Phys. Appl. 16, 353, 1981.
Chanin, M. L., The intercomparison campaign held in France in June 1981: description of the campaign, Planet. Space Sci., 31, 7, 707, 1983.
Dutsch, H. U., Vertical Ozone distribution and tropospheric ozone in Proceedings of the NATO advanced study institute on atmospheric ozone, edited by M. Nicolet and A. C. Aikin, U.S. Department of Transportation, Washington, D.C., 1979.
Elterman, L., UV, visible and IR attenuation to 50 km, Rep. AFCRL 68-1053, Cambridge Res. Lab., U.S. Air Force, Bedford, Mass., 1968.
Elterman, L., Relationships between vertical attenuation and surface meteorological range, Appl. Opt., 9, 1804, 1970.
Inn, E. C. Y., and Y. Tanaka, Absorption coefficient of ozone in the ultraviolet and visible regions, J. Opt. Soc. Am., 43, 870, 1953.
Krueger, A. J. and R. A. Minzner, A mid-latitude ozone model for the 1976 U.S. Standard Atmosphere, J. Geophys. Res., 81, 4477, 1976.
McCormick, M. P. and T. J. Swissler, Stratospheric aerosol mass and latitudinal distribution of the El Chichon eruption cloud for October 1982, Geophys. Res. Letters, 10, 0, 877, 1983.
Megie, G., J. Y. Allain, M. L. Chanin, and J. E. Blamont, Vertical profile of stratospheric ozone by lidar sounding from the ground, Nature, 270, 329, 1977.
Megie, G. and R. T. Menzies, Complementarity of UV and IR differential absorption lidar for global measurements of atmospheric species, Appl. Opt., 19, 1173, 1980.
Megie, G. and J. Pelon, Measurements of the ozone vertical distribution (0-25 km): comparison of various instruments, GAP-Observatoire de Haute Provence, June 1981, Planet. Space Sci., 39, 7, 791, 1983.
Pelon, J., P. Flamant, M. L. Chanin and G. Megie, Intrusion d'ozone d'origine polaire aux latitudes moyennes: Mise en evidence par sondage laser, C.R. Acad. Sci. Paris, 292, 319, 1981.
Pelon, J. and G. Megie, Ozone monitoring in the troposphere and lower stratosphere: evaluation and operation of a ground-based lidar station, J. Geophys. Res., 87, C7, 4947, 1982a.
Pelon, J. and G. Megie, Ozone vertical distribution and total content as monitored using a ground-based active remote sensing system, Nature, Lond., 299, 137, 1982b.
Pelon, J. and G. Megie, Lidar measurements of the vertical ozone distribution during the June 1981 intercomparison campaign GAP/OHP, Planet. Space Sci., 31, 7, 717, 1983.
Rothe, K. W., H. Walther and J. Werner, Differential absorption measurements with fixed frequency IR and UV lasers, in Optical and Laser Remote Sensing, ed. D. K. Killinger and A. Mooradian, Springer Verlag, 1983.
Schotland, R. M., The determination of the vertical profile of atmospheric gases by means of a ground-based optical radar, in Proceedings of the Third Symposium on Remote Sensing of Environment, University of Michigan, Ann Arbor, 1964.

Schotland, R. M., Errors in the lidar measurement of atmospheric gases by dif-
    ferential absorption, J. Appl. Meteorol., 13, 71, 1974.
Seinfeld, J. H., Air Pollution-Physical and Chemical Fundamentals, McGraw-Hill,
    New York, 1975.
Stewart, R. D., S. Hameed and J. Pinto, The natural and perturbed troposphere,
    IEEE Trans. Geosci. Electron., GE-16, 30, 1978.
Thompson, R. T., J. M. Hoell, and W. R. Wade, Measurements of SO_2 absorption
    coefficients using a tunable dye laser, J. Appl. Phys., 46, 3040, 1975.
Uchino, O., M. Maeda and M. Hirono, Applications of excimer lasers to laser-
    radar observations of the upper atmosphere, IEEE, J. Quant. Elec., QE-15,
    1094, 1979.
Uchino, O., M. Maeda, T. Shibata, M. Hirono and M. Fujiwara, Measurement of
    stratospheric vertical ozone distribution with a XeCl lidar: Estimated in-
    fluence of aerosols, Appl. Opt., 19, 4175, 1980.
Uchino, O., M. Maeda, H. Yamamura and M. Hirono, Observation of stratospheric
    vertical ozone distribution by a XeCl lidar, J. Geophys. Res., 88, in
    press, 1983.
Vigroux, E., Contribution experimentale a l'absorption de l'ozone, Ann. Phys.
    Paris, 12, 709, 1953.

7.   LIDAR STUDIES OF TEMPERATURE AND DENSITY USING RAYLEIGH SCATTERING

M. L. Chanin and A. Hauchecorne

Service d'Aeronomie du C.N.R.S.
Verrieres-le-Buisson, France

HISTORY

Rayleigh backscattering of a laser beam by atmospheric molecules has been observed by most of the first lidar workers, but those observations did not lead to any major results. This early work has been reviewed by KENT and WRIGHT (1970). In all these preliminary experiments, the lasers used were powerful Q-switched ruby lasers operating at a unique wavlength (694.3 nm) and the discrepancies observed with the current density model were either attributed to saturation effect (due to the high energy per pulse of the laser) or to the presence of aerosol layers. The first geophysical study using Rayleigh scattered echoes was carried out between 70 and 100 km with the very powerful lidar of Jamaica, MARK II, and indicated the existence of tidal modes in that height range (KENT and KEENLISIDE, 1975). From 1977 to 1980, a series of preliminary results was obtained from the lidar station of the Observatory of Haute Provence in France, as a by-product of the lidar sounding of sodium and temperature profiles, proving then the potential use of such a method to study the temporal and spatial fine structure in atmospheric density and temperature (HAUCHECORNE and CHANIN, 1980). In June 1980 a lidar specially devoted to that type of measurements was set up at the same site, and a systematic survey of the middle atmosphere from 30 km up to 90 km was undertaken by the authors; this station has been used on a continuous basis since June 1981 for this purpose and, as it is still the only one in operation, it will be used as a reference in this paper.

DESCRIPTION OF THE METHOD

(a) Generalities about the Lidar Technique: A lidar (acronym for light detection and ranging) is based on a principle similar to the one on which the radar operates, but in the optical spectral range. Monochromatic laser pulses are sent vertically into the atmosphere and a temporal analysis of the backscattered signal provides information about the structure and the composition of the atmosphere as a function of height; the lower limit of the height resolution is given by the laser pulse duration: 1.5 m to 150 m, for standard lasers with pulse duration ranging from 10 nsec to 1 μsec.

The backscattered wavelength could be either identical to the emitted wavelength (RAYLEIGH, MIE and resonant scattering) or situated at a higher wavelength (RAMAN and fluorescent scattering). In the problem which we are concerned with in this paper, the parameter to be measured is the molecular density of the atmosphere; the laser wavelength should then be taken outside of any absorption bands or resonance lines, and the emission and reception occur at the same wavelength. The cases where specific atoms or molecules are to be detected by resonance and absorption are treated in the same document respectively by B. CLEMESHA, and G. MEGIE and J. PELON.

When a monochromatic wavelength is sent in the atmosphere, two processes can provide a backscattered signal: the Rayleigh scattering by atmospheric molecules and the Mie scattering by atmospheric aerosols. At the height above 30 to 35 km the Mie contribution is negligible compared to the Rayleigh one, and the backscattered echo is then proportional to the atmospheric density. The temperature is deduced from the measurement of density, assuming that the atmosphere is in hydrostatic equilibrium and obeys the perfect gas law.

(b) Lidar Equation: If the backscattered light is only due to Rayleigh and Mie scattering, the lidar equation can be written:

$$N(z_i) = \frac{N_o \, A \, K \, R_q \, T^2(z_o, \, z_i)}{4\pi(z_i - z_o)^2} \cdot [n_r(z_i) \, \beta_r + n_m(z_i) \, \beta_m(z_i)] \, \Delta z \qquad (1)$$

$N(z_i)$ is the number of detected photons for one laser pulse, from a layer of thickness Δz centered at the height z_i

N_o is the number of photons emitted for each laser pulse

A the telescope area

K the optical efficiency of the lidar system (including the optical transmission through the transmitter and receiver)

R_q the quantum efficiency of the photomultiplier

$T(z_o, \, z_i)$ the atmospheric transmission between the altitude of the lidar site and the height of the emitting layer z_i

$n_r(z_i)$ and $n_m(z_i)$ the air molecules and aerosols concentrations

β_r and $\beta_m(z_i)$ the Rayleigh and Mie backscattering cross-sections.

In the height range when the Mie contribution is negligible (i.e. between 35 and 80 km as it will be shown later) the atmospheric density is given by the expression:

$$\rho(z) = C \, [S_L(z) - B(z)] \, / \, T^2(z, \infty) \qquad (2)$$

$S_L(z)$ is the signal coming from the altitude z, in a constant solid angle (i.e. multiplied by $(z - z_o)^2$) and eventually corrected for the non linearity of the photomultiplier if this one is close to saturation.

$B(z)$ is the background signal due to dark current and sky background, extrapolated linearly for the altitude z (The recovery time of a photomultiplier exposed to a high level of light may be a source of error in estimating the background coming from the dark current. To reduce such a source of error, a shutter should be used during the return of the low altitude echo; both mechanical and electronic shutters have been used successfully).

$T(z, \infty)$ is the atmospheric transmission between z and the top of the atmosphere, evaluated at the laser wavelength, taking into account ozone and Rayleigh attenuation (Handbook of Geophysics).

C a normalisation constant, depends upon N_o, $K \, R_q$ and $T(z_o, z_i)$ as defined above and may vary with time (mainly under the influence of fluctuations in the laser energy output and the atmospheric transmission). For each period of measurement C is evaluated by fitting the density measured either with a model (CIRA 1972) between 30 and 35 km, or with radio-sonde data at 30 km obtained from the nearest meteorological site.

The relative uncertainty on the density determination is given by

$$\frac{\Delta\rho}{\rho} = \frac{\Delta S_L(z)}{S_L(z) - B(z)} \qquad (3)$$

(c) Temperature Determination: The temperature profile is computed from the density profile assuming that the atmosphere obeys the perfect gas law and is in hydrostatic equilibrium. This second assumption implies that atmospheric turbulence does not affect the mean air density, which is the case considering the temporal and spatial resolutions of the lidar data. The constant mixing ratio of the major atmospheric constituents (N_2, O_2 and Ar) and the negligible value of the H_2O mixing ratio justify the choice of a constant value M for the air mean molecular weight. The air pressure $P(z)$, density $\rho(z)$ and temperature $T(z)$ are then related by:

$$P(z) = \frac{R \, \rho(z) \, T(z)}{M} \tag{4}$$

$$dP(z) = -\rho(z) \, g(z) \, dz \tag{5}$$

where R is the universal gas constant and $g(z)$ the acceleration of gravity. The combination of Equation (4) and Equation (5) leads to:

$$\frac{dP(z)}{P(z)} = \frac{M \, g(z)}{R \, T(z)} \, dz = d \, (\text{Log} \, P(z)) \tag{6}$$

If the acceleration of gravity and the temperature are assumed to be constant in the ith layer, the pressure at the bottom and top of the layer are related by:

$$\frac{P(z_i - \Delta z/2)}{P(z_i + \Delta z/2)} = \exp \frac{M \, g(z_i)}{R \, T(z_i)} \, \Delta z \tag{7}$$

and the temperature is expressed as:

$$T(z_i) = \frac{M \, g(z_i) \, \Delta z}{R \, \text{Log} \, P(z_i - \Delta z/2) \, / \, P(z_i + \Delta z/2)} \tag{8}$$

The density profile is measured up to the nth layer (about 80 km). The pressure at the top of this layer is fitted with the pressure of the CIRA 1972 model, $P_m(z_n + \Delta z/2)$, for the corresponding month and latitude. The top and bottom pressures of the ith layer are then:

$$P(z_i + \Delta z/2) = \sum_{j=i+1}^{n} \rho(z_j) \, g(z_j) \, \Delta z + P_m(z_n + \Delta z/2) \tag{9}$$

$$P(z_i - \Delta z/2) = P(z_i + \Delta z/2) + \rho(z_i) \, g(z_i) \, \Delta z \tag{10}$$

Let X be:

$$X = \frac{\rho(z_i) \, g(z_i) \, \Delta z}{P(z_i + \Delta z/2)} \tag{11}$$

The temperature is then:

$$T(z_i) = \frac{M \, g(z_i) \, \Delta z}{R \, \text{Log} \, (1 + X)} \tag{12}$$

The statistical standard error on the temperature is:

$$\frac{\delta T(z_i)}{T(z_i)} = \frac{\delta \, \text{Log} \left| 1 + X \right|}{\text{Log} \left| 1 + X \right|} = \frac{\delta X}{(1 + X) \, \text{Log} \, (1 + X)} \tag{13}$$

with

$$\left(\frac{\delta X}{X}\right)^2 = \left| \frac{\delta \rho(z_i)}{\rho(z_i)} \right|^2 + \left| \frac{\delta P(z_i + \Delta z/2)}{P(z_i + \Delta z/2)} \right|^2 \tag{14}$$

$$\delta P(z_i + \Delta z/2)^2 = \sum_{j=i+1}^{n} \left| g(z_j) \, \delta \rho(z_j) \, \Delta z \right|^2 + \left| \delta P_m(z_n + \Delta z/2) \right|^2 \tag{15}$$

The uncertainty on the extrapolated pressure at the top of the profile is evaluated to be 15%. Its contribution to the temperature uncertainty decreases rapidly with altitude and is smaller than 2%, at 15 km from the top, and smaller than 1%, 5 km lower. It is important to notice that the term X represents a ratio of experimental density values and consequently the constant of normalisation disappears. The temperature determination is then <u>absolute</u> as soon as one can neglect the term due to the pressure at the top, even though the density is only measured in a relative way.

DESCRIPTION OF THE EXPERIMENT

The instrument is made up basically from the same elements as any lidar, a transmitter and a receiver, but its specific purpose to measure density in the stratosphere and mesosphere leads to a different choice of its characteristics. We would like to suggest in order to differentiate one lidar from another to refer to the lidar described in this paper as a middle atmosphere Rayleigh lidar. Our experience in using the lidar technique in all the height range from 0 to 100 km and with a variety of interaction processes has shown that a lidar system can reach better performances if it is designed for a well defined and unique type of measurement.

The characteristics of the Rayleigh lidar in operational use at the Observatory of Haute Provence (O.H.P.) are given in Table 1 and a schematic diagram is presented in Figure 1. Arguments for the choice of the different elements, and improvements planned for the near future are discussed below.

TABLE 1

LIDAR CHARACTERISTICS

LASER

Type	Neodymium — Yag (Quantel Model 408)
Energy per pulse	400 mj at 532 nm
	(150 mj at 355 nm)
Repetition rate	10 Hz
Pulse width	15 ns
Divergence	4.10^{-4} rad
Divergence	10^{-4} rad (with a beam expander)

RECEIVER

Telescope diameter	80 cm
Telescope area	0.5 m^2
Field of view	10^{-3} to 10^{-4} rad
Band pass filter	0.8 nm (FWHH)
P.F. interferometer	20 pm (FWHH)
Gate width	4 μm (0.6 km)

(a) Transmitter: The active part or the transmitter should be a pulsed laser delivering a high average power. Rayleigh scattering cross-sections varying as λ^{-4}, the optimum choice of wavelength lies in the blue-green part of the spectrum, but the optimum choice of laser obviously depends upon the energy per pulse and repetition rate that it is able to provide for long periods of continuous operation (~ 12 hours or more). Such a choice would vary with time as a function of the state of the art, and the choice made a few years ago may become obsolete in the next few years; it may have to be revised if a new type of laser becomes available and reliable for field operation (i.e. the excimer XeF
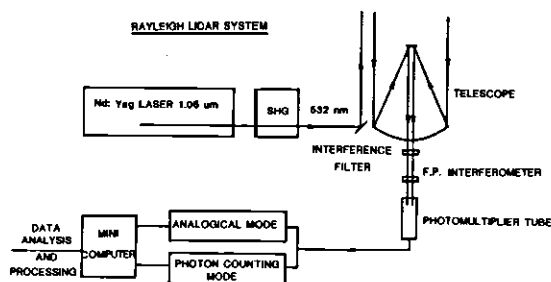
RAYLEIGH LIDAR SYSTEM

Figure 1. Schematic diagram of the present
Rayleigh lidar at OHP.

laser emitting at 353 nm). To that date, the more efficient and reliable laser
is the Nd[+++]:Yag laser with a mean power of 10 W at 1.06 μm. The second and
third harmonics of this laser emitting 5 W at 532 nm and 2 W at 355 nm re-
spectively can both be used with equivalent efficiency. It should be noticed
that at equivalent mean average power the laser with the highest energy per
pulse will be better for daytime operation. Considerations should also be
given to the natural divergence of the laser and this argument pleads for a
choice of a lower energy per pulse and higher repetition rate.

In the case of the Nd[+++] Yag laser, the natural divergence is 5 x 10^{-4}
radian which makes it possible for nightime measurements to avoid using any
transmitting optics. The beam is sent directly upward by using a total re-
flection prism.

For daytime operation both the spatial and spectral characteristics of the
laser should be upgraded. This is done by using a 10 cm collimating telescope
and inserting a Fabry-Perot interferometer in the laser cavity. The divergence
is then reduced to 1 x 10^{-4} radian and the emitted spectral width to 10 pm.

(b) Receiver: Even though the receiver is basically a light collector, it has
been proved that telescopes of good optical quality ($\sim \lambda/4$) are necessary for
optical efficiency and good rejection of sky background. Assuming this quality
is insured, then the larger telescope yields the better accuracy and the larger
range.

The results which will be mentioned in this paper were obtained with a
reflecting telescope of 80 cm diameter with a f/3 aperture. The telescope is
fixed and pointed to the zenith. The field of view can be adjusted from 10^{-3}
to 10^{-4} radian, compatible with the laser divergence and the quality of the
alignment between the transmitter and the receiver. The signal is filtered
through a narrow band filter of 0.8 nm FWHH for nighttime measurements. A
supplementary filtering, necessary for daytime measurements is made by using a
Fabry-Perot interferometer of 20 pm. The signal is detected with a Hammamatsu
R 928 photomultiplier.

It is obvious that a straightforward way of improving lidar efficiency is
to increase the size of the telescope. But, the cost of such improvement was
better used, in the last decade, by taking advantage of the rapid development in
laser technology.

(c) Electronic Processing and Data Acquisition: The major difficulty in
signal analysis comes from the large dynamical range of the backscattered

signal, which is 2 x 105 between 30 and 90 km. Up to now we accommodate such
a range by using simultaneously two acquisition modes:

- The lower altitude range (below 35 km) is analysed using a analogical sampler
  with 2048 channels and a time resolution of 100 ns corresponding to a maximum
  altitude resolution of 15 m.

- For the highest range (above 30 km) a 256 channel photon counter is used in
  parallel. The timegate of the photon counter can be varied from 1 to 8 µs,
  but has been mainly used at 4 µs providing a maximum height resolution of
  600 m.

     The data from the two acquisition systems are sent to MINC 11-02 computer
and are stored on a floppy disk. A presummation of the single-shot signals
are made over 5 minutes in order to reduce the volume of data to be stored.

     It is to be noticed that the signal coming from Rayleigh scattering at
30 km is usually low enough to be analysed in the pulse counting mode, but the
analogical sampling is useful to look at the aerosol contents in order to avoid
contamination of the data by Mie scattering. A more satisfactory approach to
handle the large dynamical range of the signal is to record the data simulta-
neously with two channels having different sensitivity. This improvement will
be implemented in the equipment in the near future.

     It is obvious that the choice of acquisition systems depends upon the
energy of the laser to be used. In the choice of a compromise between energy
by pulse and repetition rate, the performance of the acquisition system should
be considered.

DISCUSSION ON THE METHOD

     In this section, data will be presented to justify the hypothesis made
about the negligible contribution of Mie scattering and then comparison of the
lidar results with those obtained by other techniques will be discussed.

(a) The Role of Mie Scattering: In most of the early results, the differences
between density measurements and models were attributed to the presence of
aerosol layers and such layers were "seen" in both the high stratosphere and
the low mesosphere and were observed to be highly variable. Because the only
available powerful laser was at that time the ruby laser, measurements were only
performed at one wavelength, which is insufficient to attribute the observed
deviations from the model to the presence of superimposed Mie scattering. From
the data we have acquired in the last few years, it is confirmed that highly
variable structures and large difference with the models are observed, and
that they can be explained by the presence of waves (CHANIN and HAUCHECORNE,
1981; HAUCHECORNE and CHANIN, 1982-1983; CHANIN et al., 1983).

     But the presence of structures due to wave propagation does not eliminate
the possibility of detecting aerosol layers. Measurements at two different
wavelengths were performed to check about such a possibility. The different
wavelength dependence of Mie and Rayleigh scattering (roughly $\lambda^{-1.5}$ for aero-
sols and λ^{-4} for molecules) permits the identification of aerosols. The
ratio of the signals obtained at 532 nm and at 355 nm, presented in Figure 2
indicates, as expected, the presence of an aerosol layer up to 33 km, due to
the El Chichon cloud; but above that altitude the ratio of the signal at 532 nm
and 355 nm stays constant in a satisfactory way, mainly if one considers the
statistical uncertainty and the non-simultaneity of the measurements is highly
desirable (and will be available in our lidar within a month from now). It
will provide on a regular basis an estimate of the upper limit of aerosols in
the high stratosphere and mesosphere.

Furthermore, if the nature of the aerosols, and therefore the wavelength dependence of the Mie scattering cross-section, would be known, the two wavelength measurements would allow an accurate extraction of the Mie scattered component, in order to use the corrected data for density and temperature determinations in the height range where aerosols are present (i.e. below 30 km).

(b) *Comparison Between Lidar and Rocket Data:* Sounding rockets have been the only way for a number of years to obtain temperature profiles with a high vertical resolution and the series of rocket data collected from different launching sites for the last 20 years has been very precious for studying the middle atmosphere structure. However the sporadic character of the rocket launchings and the instantaneous view of the atmospheric medium it provides, prevent the detailed study of the propagation of gravity and planetary waves. Furthermore, the corrections required on the data above 60 km due to dynamical and radiative heating of the thermistors, give little faith in the temperature values in the high mesosphere. Even if limited to below 60 km, comparison between lidar and rockets data are of a definite interest and were performed on two opportunities: in 1980 a comparison with preliminary lidar data was reported (HAUCHECORNE and CHANIN, 1980) and a more recent comparison is shown in Figure 3. Agreement within ±2 K between 30 and 50 km is very satisfactory taking into account the 120 km distance between the 2 sites and the different temporal resolution of the data.
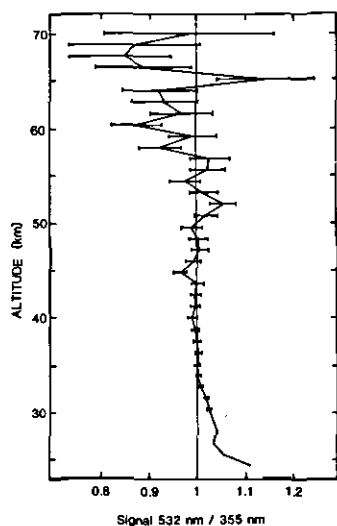
Figure 2. Ratio of the backscattered signals obtained on May 3, 1983 at 532 and 355 nm. The error bars represent ±1 standard deviation.
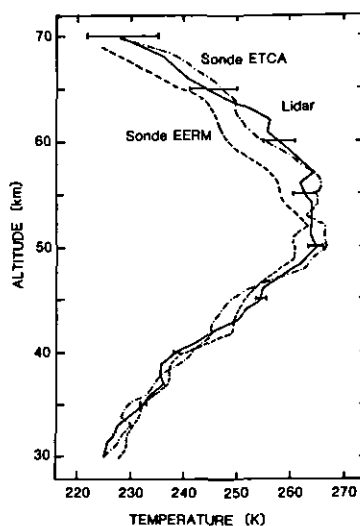
Figure 3. Vertical temperature profiles obtained on November 4, 1981 with the lidar (-) and with two different sondes on board a meteorological rocket (dashed line).

(c) *Comparison Between Lidar and Satellite Data:* Stratospheric temperatures are obtained on a global scale from CO_2 radiance at 3 different height levels (1.5, 6 and 15 mb) from the SSU experiment on board NOAA 6 and 7. During the winter 1982-1983, a comparison was performed for 60 nights of lidar data with the channel 27 of SSU. The lidar temperature profiles were completed below 30 km with the radiosonde data and for both sets of data the height resolution was

degraded to fit with the weight function of channel 27. The results of this comparison presented in Figure 4 indicate a + 2 K difference between SSU and the lidar in December 1982 but a very close agreement (~ 0.5 K) in February and March 1983. Whatever is the explanation for the difference observed in December, the general agreement is very satisfactory, mainly when considering that the measurements are not performed at the same local time and that the variability during that period was of large amplitude. Such a result is very promising for future complementary use of local high resolution lidar data and global coverage provided by satellites.
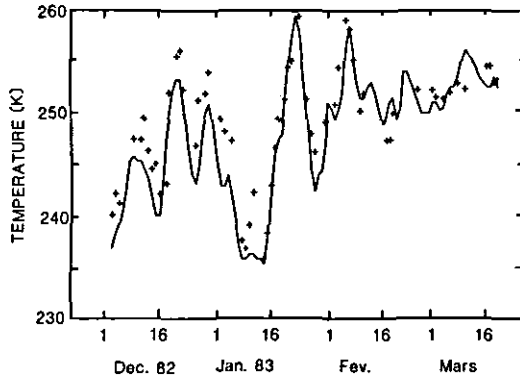

Figure 4. Brilliance temperature of SSU at 1.5 mb from channel 27 (+) compared with the lidar temperature taking into account the weight function of channel 27 (-).

PRESENT ACCURACY AND POSSIBILITY OF IMPROVEMENTS

The accuracy of the density and temperature measurements depends upon the number of photons $N(z)$ received from the height range Δz and during the time Δt; it varies as $\sqrt{N(z)}$ or as $\sqrt{\Delta t \cdot \Delta z}$. Accuracy will therefore be dependent on the time and space resolutions required and those will vary with the problem to be studied. Figure 5 illustrates the density and temperature accuracies in the altitude range 30 to 80 km for different integration time and height resolutions. The curves are obtained from the experimental results using the Rayleigh lidar described above which is in continuous use since June 1981. It represents the averaged performances of a routinely run instrument and not the exceptional values which could be obtained sporadically. Extrapolation of these values for lidars having different characteristics are deduced easily from these curves.

Since our first measurements in 1977, a series of successive improvements have increased both the range and accuracy of the measurements. More improvements are expected in the future: reduction of the background noise, increase of the dynamic range, and simultaneous use of two wavelengths. But the major improvements would come from increasing the average laser power (mainly if high repetition rate lasers become reliable) and/or from increasing the size of the telescope. It is expected in the future that the range of a Rayleigh lidar could reach the 100 km level with accuracy identical to the ones observed today at 80 km.

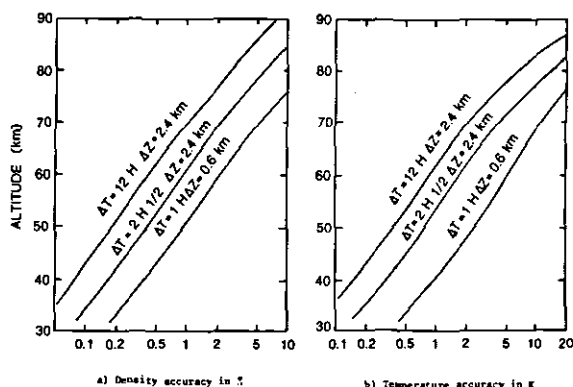Another major improvement would come from the possibility of tilting the

Figure 5. Accuracy in the density (a) and
in the temperature (b) measured from
Rayleigh lidar with different time and
height resolutions.

telescope axis in order to get a 3-dimensional view of the medium, and measure
the horizontal extend of small scale structures.

With the performances already available, the advantages of a Rayleigh
lidar, in terms of atmospheric needs, are mainly due to three characteristics:

- The possibility of making absolute temperature determination without requir-
  ing an external calibration, and this with an accuracy better than 1 K in the
  stratosphere.

- The availability of good vertical resolution (~ 150 m) and, if the system is
  used in a scanning mode, a high horizontal resolution (~ 1 m).

- The continuity of the survey in altitude, from 30 to 80 km, and in time. The
  observations are only limited by meteorological conditions (which implies a
  good choice of the site) and now to nighttime (but this limit should soon
  disappear).

These three main characteristics make the Rayleigh lidar a powerful tool
mostly for two types of studies based either on the knowledge of the absolute
temperature and its eventual evolution, or on the description of the temporal
and spatial fine structure of both density and temperature. For both ap-
proaches it will be a necessary complement to rockets and satellite observa-
tions.

EXPERIMENTAL RESULTS ALREADY ACHIEVED

Typical density and temperature profiles measured by the Rayleigh lidar
method are represented in Figure 6. Data acquired sporadically during the
years 1977-1980 and on a regular basis since 1981 have already led to a number
of geophysical results which will be briefly summarized.

(a) Long Term Variation of Stratospheric and Mesospheric Temperature: Lidar
measurements are able to provide a survey of the stratospheric temperature with
an accuracy better than 1 K which should be adequate for studying the influence
of the 11-year solar cycle or detecting the effect of long term perturbations
from volcanic or anthropogenic origins. The series of observations is yet too
short to lead to definite conclusions but already indicates variation which are
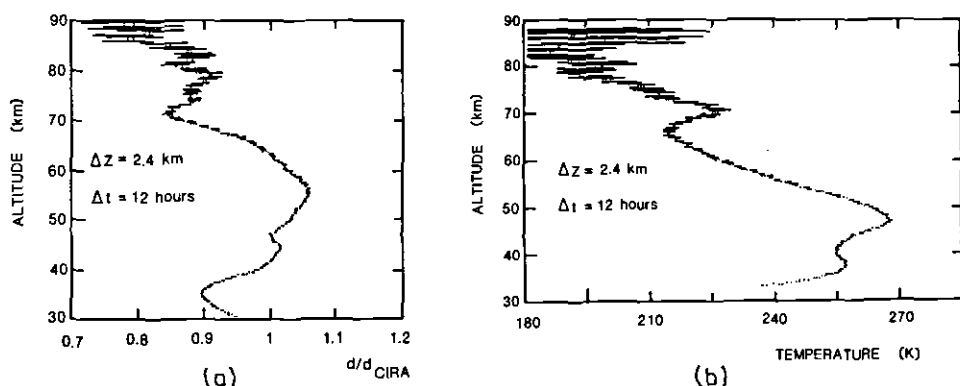
Figure 6. Density (a) and temperature (b) profiles from lidar
for the night of January 19, 1983 (the error bar represents
± one standard deviation). The density is referred to the
density from the CIRA 1972 model.

above the detection level. Correlated study with long term variations of ozone
and aerosols measured by other means are also undertaken in order to identify
a possible relationship between the variations observed.

(b) Validation of Satellite Temperature Data: The absolute determination of
temperature can be helpful to calibrate the radiometers placed onboard opera-
tional satellites as was done with SSU data, in order to detect an eventual
drift and to relate the results from one satellite experiment to the following
one. The high vertical resolution of the temperature profile has also been
used to aid the inversion procedure of satellite radiance data mainly during
very disturbed periods. Such a comparison of a lidar profile and a SAMS re-
trieved temperature profile has been done in several opportunities with suc-
cess.

(c) Contribution to a Middle Atmosphere Reference Atmospheric Model: The
quantity of data acquired during the last years (1980-1983) has already been
used to study the seasonal variation and the year-to-year variability. Notice-
able differences with the CIRA 1972 model were observed but they may only be
characteristic of the region (44°N, 6°E) where they were observed. On the
other hand, the seasonal variability of the day-to-day temperature variance
was shown to present very large and reproducible features and it is suggested
that the notion of variance will be introduced in the next CIRA model (BOURG-
HECKLY et al., 1982).

(d) Wave Studies: The Rayleigh lidar provides a continuous measurement from
30 to 80 km of the density and temperature with a height resolution of 600 m (to
be reduced soon to 150 m). Profiles can be obtained for an integration time of
half an hour with a reasonable accuracy (Figure 7) and this continuously during
the whole night. Then it becomes possible to study the temporal and spatial
characteristics of gravity waves of periods larger than one hour and of verti-
cal wavelength larger than 2 km (waves with such characteristics are responsi-
ble for the transport of a large part of the energy up to the mesosphere). If
the profiles are integrated over several hours, the structures due to gravity
waves disappear, and such profiles, if obtained regularly, can be used to study
planetary waves. Furthermore the extension of the observations to daytime will
permit 24 hours of continuous observations and the study of the diurnal and
semi-diurnal tides. Results have already been obtained on both planetary waves
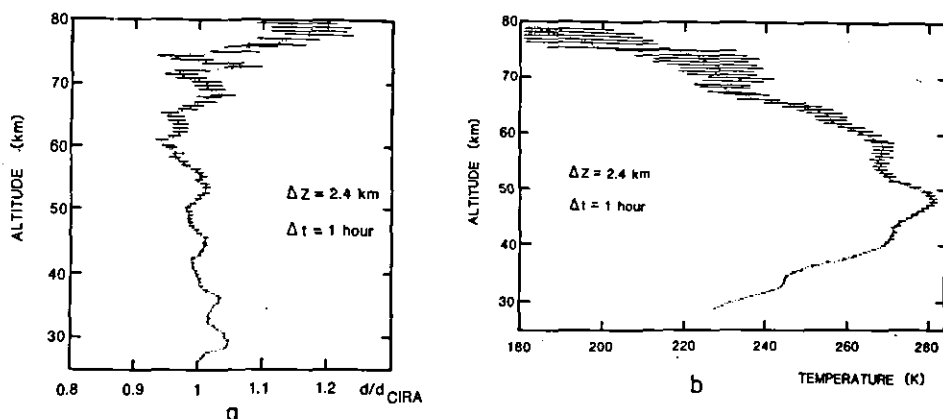and gravity waves and will be briefly summarized.

Figure 7. Density (a) and temperature (b) profiles from lidar
for one hour integration time on the night on June 10, 1983.

Regular survey of the temperature during the last three winters, as shown
in Figure 8, have demonstrated the simultaneous presence of Rossby waves of
periods varying from 5 to 20 days and vacillations of the zonal circulation of
periods 25 to 60 days. Such waves are shown to be present in the high strato-
sphere and the mesosphere continuously during the winter (from October to March)
and are responsible for the succession of observed stratospheric warmings and
mesospheric coolings. The high vertical resolution in the high stratosphere
and mesosphere, where the maximum amplitude of these effects are observed, pro-
vides a possible diagnostic to predict the occurrence of stratospheric warmings
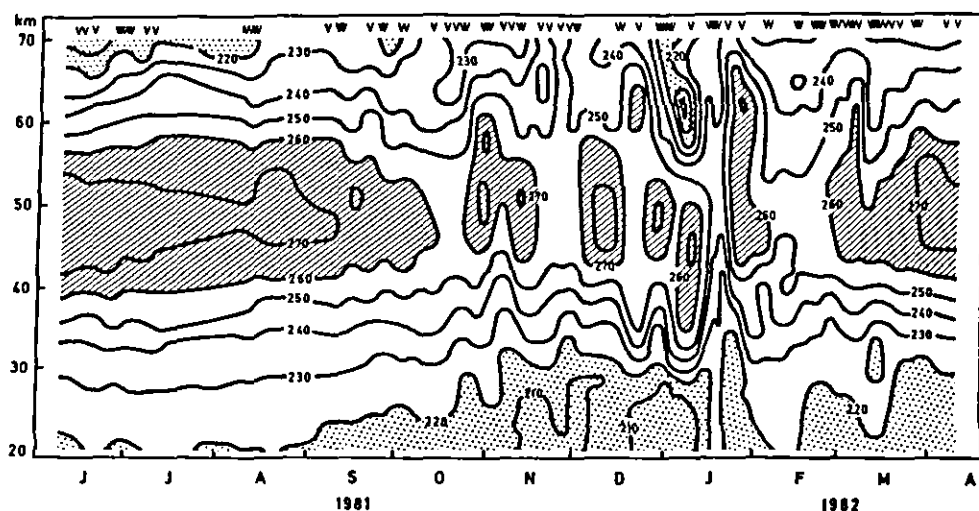(HAUCHECORNE, 1983).

Figure 8. Isocontours of temperature obtained by lidar during
the period June 1981-April 1982. The arrows at the top indicate
days when data were taken.

The study of gravity waves are in a preliminary stage but two types of waves have been observed: Lee waves of about 5 km vertical wavelength and waves of about 10 km ± 3 vertical wavelength, propagating upward with a vertical phase speed of 0.1 to 1 ms^{-1}. Breaking of the waves when they reach saturation ($\frac{\delta T}{\delta z} \sim \Gamma$) is seen to occur often at altitudes as low as 50 km and the behavior of the wave amplitude around the breaking level has been studied in detail. Only preliminary results have been published (CHANIN et al., 1983). But it already appears that the contribution of the Rayleigh lidar is complementary to the MST and partial-reflection radar's contribution. The simultaneous use of both techniques on the same site will allow the evaluation of both momentum fluxes and heat fluxes.

CONCLUSION

The instrument described in this paper has now proved to be a necessary complement to the different tools already available to study the atmosphere either to complete the altitude range to be studied or to measure a complementary parameter. Furthermore, the high spatial resolution provides a new insight in the behavior of the middle atmosphere. However it is obvious that the results have to be used in correlation with satellite observations in order to be placed in the global context. A major contribution to the understanding of the dynamics of such instruments, mainly if they are combined with radars, meteorological rockets and balloons. The spacing of such a network and the geographic situation of the sites depend on the problem to be studied.

REFERENCES

Bourg-Heckly, G., A. Hauchecorne and M. L. Chanin (1982), Variation saisonniere de la temperature de l'atmosphere a moyenne latitude, Contribution a la revision du modele CIRA 1972, Notes C.R. Acad. Sci., Paris, 295, 997-1002.

Chanin, M. L. and A. Hauchecorne (1981), Lidar observation of gravity and tidal waves in the middle atmosphere, J. Geophys. Res., 86, 9715-9721.

Hauchecorne, A. and M. L. Chanin (1980), Density and temperature profiles obtained by lidar between 35 and 70 km, Geophys. Res. Letts., 7, 565-568.

Hauchecorne, A. and M. L. Chanin (1982), A mid-latitude ground-based study of stratospheric warmings and planetary wave propagation, J. Atmos. Terr. Phys., 44, 577-589.

Hauchecorne, A. and M. L. Chanin (1983), Mid latitude lidar observations of planetary waves in the middle atmosphere during the winter of 1981-1982, J. Geophys. Res., 88, 3843-3849.

Hauchecorne, A. (1983), Observation par lidar et modelisation des echauffements brusques et des ondes planetaires dans l'atmosphere moyenne, Doctorate Thesis, University of Paris.

Kent, G. S. and R. W. H. Wright (1970), A review of laser radar measurements of atmospheric properties, J. Atmos. Terr. Phys., 32, 917-943.

Kent, G. S. and W. Keenslide, (1975), Laser radar observations of the diurnal atmospheric tidal mode above Kingston, Jamaica, J. Atmos. Sci., 32, 1663-1666.

8. LIDAR STUDIES OF THE ALKALI METALS

B. R. Clemesha
Instituto de Pesquisas Espaciais - INPE
Conselho Nacional de Desenvolvimento Cientifico e Tecnologico - CNPq
C.P. 515, 12200 - Sao Jose dos Campos, SP, Brasil

ABSTRACT

The development of the lidar technique in the early sixties, and the sub-
sequent introduction of tunable lasers, made accurate measurements of the ver-
tical distribution of the alkali metals in the atmosphere possible for the
first time. Over the last decade a great deal of information has been obtained
on the spatial and temporal variations of sodium, and rather less information
has been obtained about potassium and lithium. The possibility of making con-
tinuous observations of the vertical distribution of sodium, coupled with
temperature measurements via the determination of the Doppler spectrum of the
returned lidar signal, offers a potentially useful technique for studying the
dynamics of the 80 to 100 km region of the atmosphere.

INTRODUCTION

Efforts to probe the atmosphere by active optical techniques started be-
fore World War II. SYNGE (1930) first suggested that scattering of light from
a searchlight beam could be used to determine atmospheric densities, and HULBURT
(1937) appears to have been the first to successfully apply the technique and
made measurements at heights up to 65 km.

The main limitation of the searchlight technique was that the light source
was continuous, so that height resolution could be achieved only by separating
the searchlight and the receiver by many kilometres, and defining a scattering
volume by the intersection of the searchlight beam and the cone of sensitivity
of the receiver. Early attempts to use a pulsed light source, employing dis-
charge tubes (FRIEDLAND et al., 1956), suffered from the limitations of the
sources available. The development of the first pulsed laser by MAIMAN (1960)
solved this problem, providing a source which in may ways is ideal for an opti-
cal radar. By allowing the use of narrow bandwidth filters and a narrow beam
divergence, the narrow bandwidth and coherence of the laser made it possible
to reduce the receiver background noise by many orders of magnitude.

The first measurements of atmospheric parameters using laser radar, or
lidar, were made by FIOCCO and SMULLIN (1963), followed by COLLIS and LIGDA
(1966), MCCORMICK et al. (1966), BAIN and SANDFORD (1966) and CLEMESHA et al
(1966). These early measurements all used ruby lasers and were concerned with
measuring either Rayleigh scattering from the major molecular constituents of
the atmosphere, with the aim of determining atmospheric density, or Mie scat-
tering from aerosols. The development of the pulsed dye laser, tunable over a
range of wavelengths opened up the possibility of measuring the vertical dis-
tributions of selected minor constituents for which convenient resonant transi-
tions exist. This technique was first applied by BOWMAN et al. (1969), who
used it to measure the vertical distribution of atmospheric sodium.

THE BASIC LIDAR TECHNIQUE

All the lidars which have been used for studies of the alkali metals have
been monostatic, i.e. with the transmitter and receiver effectively coincident,
and there appear to be no advantages to be gained by using a bistatic system.
For this reason only the monostatic system will be discussed here. The laser
is assumed to transmit a pulse whose duration is short in relation to the re-
quired height resolution. It is easy to show (KENT et al., 1967) that the sig-

nal return, S(z), from a height, z, for such a lidar is

$$S(z) = \frac{\sigma(z)}{z^2} A \, \Delta z \, N_0 \, Q \, T(z)^2 \tag{1}$$

where $S(z)$ is the signal received from the range interval $z - \Delta z/2$ to $z + \Delta z/2$,

> $\sigma(z) = n(z) \, \rho$ is the volume backscattering coefficient at range z,
> $n(z)$ is the number density of scatterers of backscattering coefficient ρ at range z,
> A is the area of the lidar receiver,
> N_0 is the energy contained in the transmitted pulse,
> Q is the overall efficiency of the lidar receiver,
> $T(z)$ is the atmospheric transmission between the lidar and range z.

In the case of lidars designed to measure resonant scattering from the alkali metal layers, the received signal is invariably so weak that photon counting techniques must be used to measure it. In this case it is convenient to express $S(z)$ and N_0 in terms of numbers of photons.

In practice it is virtually impossible to relate the received signal to the number density of the scatterers in an absolute sense because $T(z)$, and even Q, are not accurately known. This problem is conveniently overcome, in the case of resonant scattering from the alkali metal layers, by comparing the resonant scattering with Rayleigh scattering from the main atmospheric constituents at much lower heights, where the atmospheric density is accurately known. In this way we can eliminate Q and N_0 from equation 1. If we take the case of sodium, for example, the backscattering coefficient per molecule for a typical laser bandwidth of 10 pm is about 1.4×10^{-17} m^2 SR^{-1}, and the density at the peak of the layer around 90 km is of the order of 3×10^9 m^{-3} giving a volume backscattering coefficient of about 4×10^{-8} m^{-1}, SR^{-1}. A similar volume backscattering coefficient for Rayleigh scattering from the principal atmospheric constituents is obtained at about 23 km, where the number density is about 10^{24} m^{-3}, and the effective backscattering coefficient per molecule is 3.93×10^{-32} m^{-2}, SR^{-1}. The signal from this height would be about 15 times the sodium signal because of the $1/z^2$ factor in the lidar equation. In practice a somewhat greater height would normally be used in order to avoid scattering from stratospheric aerosols.

Under normal conditions atmospheric extinction for visible wavelengths is negligible above 20 km (neglecting, for the moment, extinction by the resonant scatterers themselves), allowing us to eliminate $T(z)$ from equation 1, as well as Q and N. We can then write

$$\frac{S_1}{S_2} = \frac{z_2^2}{z_1^2} \frac{\sigma_1}{\sigma_2} \frac{\Delta z_1}{\Delta z_2}$$

where the subscript 1 refers to the alkali metal layer, and 2 refers to the stratospheric reference signal. Thus we get the alkali metal density, n_1, at height z_1 in terms of the atmospheric number density, n_2, at some reference height, z_2:

$$n_1 = \frac{S_1}{S_2} \frac{z_1^2}{z_2^2} \frac{\Delta z_2}{\Delta z_1} \frac{\rho_2}{\rho_1} n_2 \tag{2}$$

In equation 2, ρ_2, the average backscattering coefficient per air molecule, is accurately known. The determination of ρ_1, the effective backscattering coefficient per alkali metal atom, is more difficult because it depends on the laser emission spectrum. The parameter ρ_1 is given by

$$\rho_1 = \int_0^\infty I(\lambda)\rho(\lambda)d\lambda \Big/ \int_0^\infty I(\lambda)d\lambda$$

where $I(\lambda)$ is the laser energy and $\rho(\lambda)$ is the alkali atom backscattering co-

efficient, both as functions of wavelength, λ. $\rho(\lambda)$ can be calculated from a knowledge of the hyperfine structure of the resonance involved, the oscillator strengths and the temperature of the atoms. Since a reasonable estimate of the temperature can be made, there is no difficulty in determining $\rho(\lambda)$. An accurate determination of $I(\lambda)$, on the other hand, is not easy, and the laser spectrum will tend to change as a function of several laser parameters such as input energy, dye lifetime and flashtube wear. Most workers have made a spectroscopic determination of the laser bandwidth, and have calculated ρ_1 on the assumption of a Gaussian line shape. An alternative approach is to measure ρ_1 directly by sweeping the output by the appropriate alkali atoms in a scattering cell. This method gives good results, but requires a laser whose output wavelength can be shifted in small, accurately known steps.

PRACTICAL DETAILS

The major difficulty with any lidar for upper atmospheric studies is one of obtaining signal returns strong enough for an accurate determination of the required parameter to be made. At least in the case of nighttime operation, external noise due to extraneous light sources is generally less important than the shot noise in the signal being measured. Thermal noise in the receiver can usually be reduced to negligible levels by cooling the detector. On the assumption that the detected photon pulses follow a Poisson distribution, the fractional precision in the measurement is simply $1/\sqrt{S}$, where S is the total number of photons detected from a given range interval; or, more precisely, 63% of samples will fall within the range $\bar{S} \pm \sqrt{S}$ where $\bar{S}$ is the true mean value of S. If external noise is not negligible then the precision becomes $\sqrt{S}/(S-N)$, where N is the external noise.

As can be seen from equation 1, S can be increased by increasing N_0, the number of photons transmitted, proportional to the pulse energy of the laser, A, the receiver area, or Q, the receiver efficiency. The last of these parameters depends on the quantum efficiency of the photomultiplier detector used, but has a maximum value of about 20% for the sodium wavelength, and is less at the wavelengths appropriate for lithium and potassium.

Since we can accumulate the signal from pulse to pulse, N_0 is simply the total number of photons transmitted during a measurement. This means that the mean power of the laser, rather than the pulse energy, is the important parameter in this respect. On the other hand, N_0 should be large enough for the condition S >> N to hold. The pulse energy and mean power of the laser and the receiver area are major items in the cost of the lidar, and their relative values must be chosen on the basis of an incremental cost analysis. Typical values in use are 50 mJ, 50 mW, and 0.5 m^2, although much larger values can be achieved, particularly for the laser parameters, and have been used by some workers.

A further factor which must be taken into account in optimizing the lidar characteristics is the laser bandwidth. For atmospheric sodium measurements a full bandwidth of about 0.07 Å has generally been used. This is appreciably larger than the Doppler width of the line at mesopause temperatures, which is about 0.03 Å. A reduction in the laser linewidth would consequently increase the effective scattering cross-section and the received signal. On the other hand, a reduction in the laser bandwidth generally involves a reduction in output energy, and a bandwidth of 0.07 Å probably represents close to the best compromise between bandwidth and energy. It should also be remembered that narrower bandwidths mandate better laser stability, and greater difficulty in determining the effective resonant scattering cross-section.

So far we have been discussing nighttime observations of sodium where it is comparatively easy to achieve a signal much stronger than the combined sky

noise and PMT dark current. In the case of the other alkali metals, whose concentrations are much less than that of sodium, and in the case of daytime measurements, this is not always so. In this case it is necessary to use higher pulse energies and to minimize extraneous noise sources. For daytime measurements the limiting factor is sky noise, and this is minimized by the use of a narrow angular beamwidth and narrow receiver bandwidth. Beamwidths used for daytime observations are of the order of 0.2 mR, placing limitations of the sort of optics used. As an economical way of achieving large receiver areas, some lidar groups have used plastic Fresnel lenses or searchlight mirrors, but such systems cannot produce the narrow beamwidths required for daytime measurements. Narrow receiver bandwidth is achieved by inserting a Fabry-Perot interferometer in the optical path, giving a bandwidth of the order of 0.2 Å. In the case of sodium, decreasing the receiver bandwidth results in a disproportionately large decrease in noise because of the existence of a strong solar Fraunhoffer line. It should be noted that a narrow angular beamwidth is necessary, not only to reduce the area of sky viewed by the receiver, but also in order to make it possible to use the narrow bandwidth Fabry Perot interferometer.

EXTINCTION AND SATURATION EFFECTS

In the analysis presented above it has been assumed that virtually all the alkali atoms to be detected are in the ground state, and that these atoms cause negligible extinction of the lidar beam. Neither of these assumptions is entirely valid, and under certain circumstances, saturation and extinction must be allowed for.

Extinction is appreciable only in the case of sodium; for the other alkali metals the concentration is too small for the effect to be important. Even in the case of sodium the effect is significant only for exceptionally high abundances. In correcting for extinction it must be remembered that the spectrum of the lidar pulse changes as it propagates through the layer, and that the spectrum of the returned signal can be quite different from that of the transmitted one. An inadequate appreciation of these points appears to have led to a certain amount of confusion in the literature. A correct analysis of the effects of extinction has been published by Simonich and Clemesha (1983), who have shown that the magnitude of the effect depends on the bandwidth of the laser emission, and that it is greatest for a bandwidth much less than that of the atmospheric sodium. The effect of extinction is about 3% for the return from the peak of a typical layer with 5×10^{13} m^{-2} abundance. This would increase to about 5% for a narrow band laser. In practice, an iterative technique can be used to correct for this extinction.

Apart from the minor influence of the laser bandwidth, resonant extinction depends only on the scattering cross-section and number density of the alkali atoms involved. Saturation effects, in contrast, are independent of the number density of scattering atoms, but depend on the power density in the transmitted beam and the lifetime of the excited state. The saturation problem has been investigated in some detail by MEGIE et al (1978), who showed that the importance of the effect depends on the ratio of the excitation rate, proportional to the product of the power density incident on the scattering atoms and their effective scattering cross-section, to the transition probability for spontaneous emission (the Einstein A coefficient). Megie and his coworkers used a dye laser pumped by a Q-switched ruby laser to measure atmospheric potassium. The short (30 nS) pulse duration of their laser combined with a fairly narrow transmitted beamwidth (0.5 mR) resulted in a power density of about 1 MW m^{-2} at the height of the layer, causing about 10% of the potassium atoms to be in the excited state, with a consequent 10% decrease in the expected signal. MEGIE et al. point out that this effect could become important in the case of future measurements using higher powers and narrower beamwidths. On the other

hand, it seems that this difficulty, should it arise, could be overcome by using a laser with a longer pulse duration and consequently lower peak power.

THE INPE LIDAR

As an example of a lidar system in use for studying atmospheric sodium, the following is a description of the INPE lidar. This system is used to provide both sodium and stratospheric aerosol measurements simultaneously. Figure 1 is a block diagram of the system, and Figure 2 shows the optical system. More detailed views of the transmitting and receiving systems are shown in Figures 3 and 4 respectively. The transmitter uses a flashlamp pumped dye laser to give a maximum output of about 70 mJ in a 2 μS pulse. Three Fabry-Perot etalons are used to tune the output to 5890 $\overset{\circ}{A}$ and reduce the bandwidth to about 0.07 $\overset{\circ}{A}$. One of these etalons is piezo-electrically tuned and forms part of a servo loop to keep the laser tuned to the D_2 line emitted by a sodium lamp (CLEMESHA et al, 1975). Referring to Figure 3, the laser output beam is collimated by L_1, L_2, and M_1 to give a final angular beamwidth of about 0.1 mR. A small fraction of the laser output is sampled by a partially reflecting mirror, M_2, and is focussed onto a small opening in an integrating sphere. A photodiode, illuminated by the light which escapes from a second small hole in the integrating sphere, provides a signal proportional to the total laser energy, and third opening allows a small fraction of the light to enter a sodium vapor cell maintained at 106°C. The light scattered by the sodium vapor, measured by a phtomultiplier, provides a measure of the laser energy within the D_2 line width. Both the total energy and the sodium vapor scattered signal are recorded along with the lidar signal, making it possible to correct for changes in the laser output spectrum which occur during observations. It should be mentioned that such changes, resulting from aging of the dye and flashlamp, and accidental mis-alignments, are by no means negligible. A helium-Neon laser, the beam from which can be introduced into the system by means of a movable mirror (not shown in the figures) is used for alignment purposes. This monitoring system, together with the ability to tune the laser via the piezo-electrically controlled etalon, also makes it possible to determine the effective sodium scattering cross-section experimentally, as described in the section headed "Basic lidar technique".
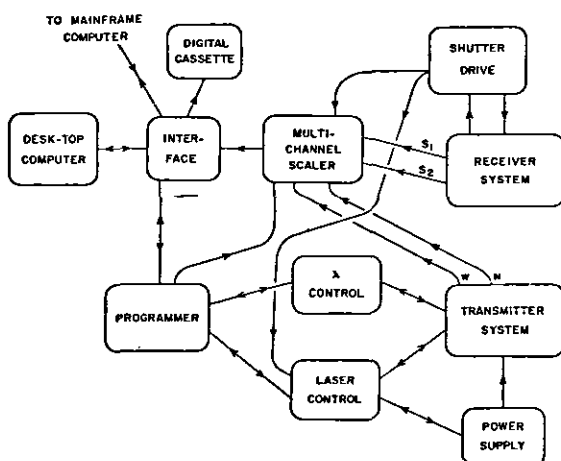

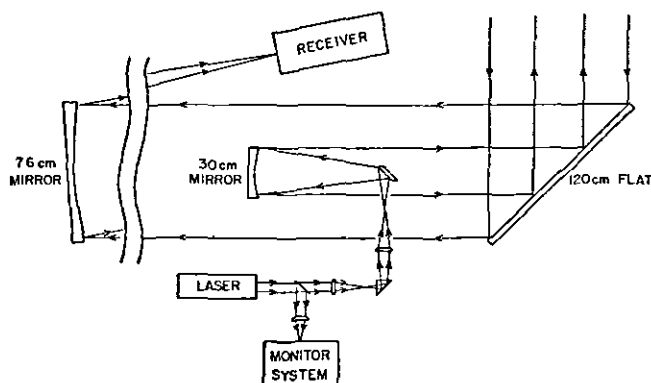Figure 1. Block diagram of the INPE lidar.

104

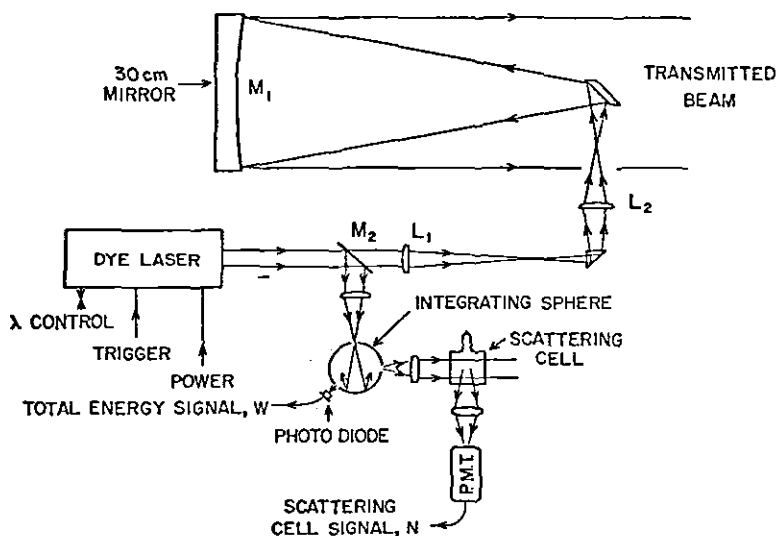Figure 2. Basic optical system of the INPE lidar.

Figure 3. Lidar transmitter.

The receiving optics use a 76 cm diameter spherical mirror of 17 m focal
length. This mirror was used because it happened to be available, and a much
shorter focal length would normally be employed. Nevertheless, there are ad-
vantages to using such a long focal length. Apart from cost considerations,
the long focal length makes it possible to use the mirror off axis without
introducing excessive aberrations. After passing through an adjustable dia-
phragm, which defines the receiver angular beamwidth, the light is collimated
in a 13 mm diameter beam which is filtered by a 7 Å bandwidth interference
filter before being focussed onto the input end of a light-guide. A beam
splitting device in the light-guide directs 95% of the light to the high sensi-
tivity photomultiplier, and the remaining 5% to the low sensitivity tube. Mul-
tiple reflection on the photocathode of the PMT is used to increase its quantum
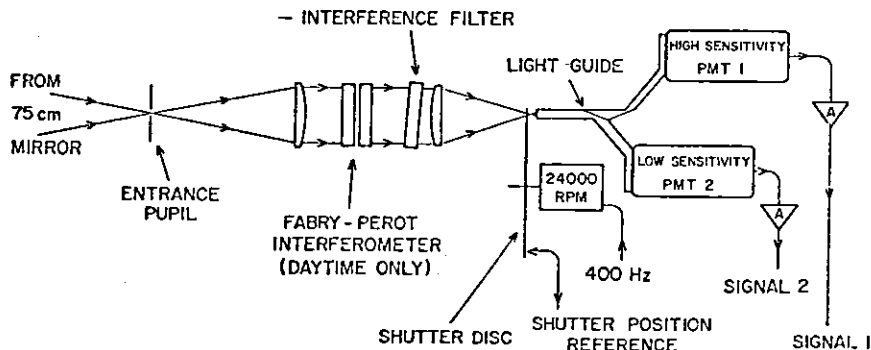
Figure 4. Receiver optics.

efficiency. The use of 2 photomultipliers in this way increases the dynamic range of the system, making it possible to record the scattering from as low as 10 km, simultaneously with the signal from the sodium layer.

Just in front of the entrance to the light guide is a rotating shutter which prevents the strong signal scattered in the lower atmosphere from reaching the PMT. For daytime observations, the receiver bandwidth is reduced by inserting a Fabry-Perot interferometer between the collimator and the interference filter.

As shown in Figure 2, the transmitting and receiving optics, which are almost coaxial, point horizontally at a 120 cm diameter plane mirror mounted on a surplus 60 inch searchlight base. In this way a limited degree of beam steering is possible, a feature which has been exploited to make spaced observations of both sodium and stratospheric aerosols (CLEMESHA et al, 1980, 1981a).

Referring to Figure 1, the pulse outputs from the photomultipliers, after amplification and threshold discrimination, are fed to a 100 channel high speed scaler, 25 channels of which are dedicated to the low sensitivity PMT, and 71 to the high sensitivity tube. The remaining 4 channels are used to record the laser energy, sodium vapor cell output, number of shots fired and time. The opening and closing of each channel is programmed in sequence, and is normally set to cover the height ranges of interest in intervals of 1 km. An accurate determination of the ratio between the sensitivities of the two photomultipliers is assured by a 16 channel overlap. The signals for time intervals corresponding to ranges from which no appreciable scattering should be detected are recorded to give values of the noise level to be subtracted from the measured photon counts. This system is somewhat antiquated (it was designed in 1968), and it will be replaced by a more flexible microprocessor-based system in the near future.

The photon counts, typicaly accumulated for 50 laser shots, are transferred to a small desk top computer which carries out a partial analysis of the data in order to monitor the system performance and the behavior of the sodium layer. The data is also recorded on digital data cassettes for subsequent transmission to INPE's mainframe computer. Laser firing and readout of the multichannel scaler is controlled by a programmer, and suitable arrangements are made to ensure that the laser firing is synchronised to the receiver shutter.

Specifications for the INPE lidar are given in Table 1.

Table 1. Specifications for the INPE lidar.

	NIGHT VALUE	DAY VALUE
Transmitted energy	30 mJ	60 mJ
Pulse duration	2 µs	2 µs
Repetition rate	$0.4 \ s^{-1}$	$0.4 \ s^{-1}$
Wavelength	5890 Å	5890 Å
Total transmitted bandwidth	0.08 Å	0.08 Å
Receiver area	$0.39 \ m^2$	$0.39 \ m^2$
Receiver bandwidth	7 Å	0.3 Å
Transmitter beamwidth	0.15 mR	0.15 mR
Receiver beamwidth	0.4 mR	0.2 mR
Receiver efficiency	2.4%	1%
Height interval	1 km	1 km

PUBLISHED RESULTS

It is not the purpose of this article to review the results of lidar studies of the alkali metal layers, but some representative results are outlined here in order to give examples of the sort of information obtainable.

Results of extensive observations of the sodium layer have been published by GIBSON and SANDFORD (1971), MEGIE and BLAMONT (1977) and SIMONICH et al (1979). Apart from a single measurement of the total abundance of potassium, published by FELIX et al. (1973), the only results of observations of potassium and lithium have been presented by MEGIE et al. (1978) and JEGOU et al. (1980) respectively. More recently the French workers have also measured calcium and ionized calcium (CHANIN, personal communication, 1983) but the results of these observations have not yet been published. The theory of the alkali metals in the atmosphere will not be discussed here. The interested reader is referred to KIRCHHOFF (1983), KIRCHHOFF and CLEMESHA (1983) and KIRCHHOFF et al (1981), and the references therein.

The following brief survey of results is divided into Vertical Distribution, Diurnal Variations, Seasonl Variations, Dynamical Effects and Temperature Measurements.

VERTICAL DISTRIBUTION

In Figure 5 we show average distributions for sodium, potassium and lithium. The curve labelled "Na Winkfield" is an annual average derived from measurements made by GIBSON and SANDFORD (1971) at a latitude of 51°N, and "Na S.J. dos Campos", for 23°S, is from SIMONICH et al. (1979). The curves of potassium and lithium, derived from data published by MEGIE et al. (1978) and JEGOU et al. (1980) respectively, are both for a latitude of 44°N. The sodium distributions shown in Figure 3 are based on a fairly large number of individual profiles taken over a complete year, and provide a good estimate of the true average vertical distribution. Considerably less data are available for potassium and lithium, and, particularly in the case of lithium, the profiles shown must be considered representative rather than truly average. As can be seen from the figure, the peak of the sodium layer is about 3 km higher at 23°S than at 51°N, and the mid-latitude profile shows significantly more sodium on the bottomside as compared with the low latitude one. This increased bottomside sodium results mainly from the large winter increase observed at mid latitudes. It is difficult to base any conclusions on the relative distributions of potassium and sodium, but, according to MEGIE et al. (1978), the peak of the potas-
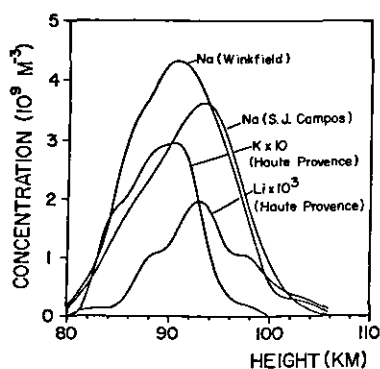
Figure 5. Average vertical distributions
of sodium, potassium and lithium.

sium layer is generally about 1 km lower than that of the sodium layer.

DIURNAL VARIATIONS

Although the first daytime lidar measurments of atmospheric sodium were
made by GIBSON and SANDFORD (1972), studies of the 24 hour variations have been
published only by CLEMESHA et al. (1982) and GRANIER and MEGIE (1982). The
French workers state that they observed "no regular variation of the character-
istic parameters of the sodium layer . . . . . during the diurnal cycle", al-
though they do not give any indication of the extent of the data set on which
they base this conclusion. The Brazilian group, in contrast, observed strong
diurnal and semi-diurnal oscillations in the layer which they observe at 23°S
(Figure 6). The oscillations appear to be mainly the result of tidally induced
vertical motions in the atmosphere, and their study would appear to provide a
useful technique for investigating atmospheric tides in this height region. A
distinctive feature of such observations is that they provide information on
the vertical tidal motions of a neutral species, unobtainable by other tech-
niques.

One important result of the daytime measurements is that they have shown
that there is no large variation between the daytime and nighttime abundances
of sodium. Earlier dayglow measurements by BLAMONT and DONAHUE (1964) had
suggested the existence of a large increase in sodium during the day, but the
more recent observation of the lack of any such enhancement has made it possible
to eliminate a number of models which had been suggested for the sodium layer
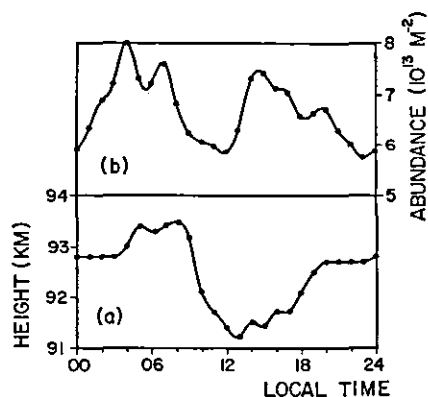(KIRCHHOFF and CLEMESHA, 1983, CLEMESHA et al., 1981b).

Figure 6. Diurnal variations of sodium
measured at Sao Jose dos Campos.

No daytime observations of the other alkali metals have been reported as yet.

SEASONAL VARIATIONS

The available data on seasonal variations is summarized in Figure 7. Sodium shows a winter maximum at all latitudes, and the main variation is clearly seasonal rather then annual. The details of the variation are not the same at different latitudes, the mid-latitude data showing a sharp increase in October, November and December, with comparatively little variation during the rest of the year, while the southern hemisphere low latitude results show a broad winter maximum, with minimum abundances in December and January. A further difference is that the winter maximum at mid-latitudes seems to be mainly the result of increased sodium on the bottomside of the layer, while at low latitudes there is no systematic annual variation in the vertical distribution.

The rather sparse lidar data available for potassium (MEGIE et al., 1977) shows no regular seasonal variation, in agreement with earlier twilight photometry results (HUNTEN, 1967). According to JEGOU et al. (1980), lithium shows an annual variation similar to that of sodium, but, in view of the paucity of the lithium data, this conclusion must be viewed with some caution.

As yet there is no satisfactory explanation for the seasonal variations. MEGIE et al. (1978) have suggested that the sharp winter peak in sodium might be caused by vertical transport of sodium compounds from the surface, and a number of workers have suggested that photochemical effects might be involved. Unfortunately, neither our knowledge of vertical transport in the atmosphere, nor our understanding of the photochemistry of sodium are sufficient for us to adequately test these ideas. Other possibilities, such as changes in the rate of sublimation of sodium from mesospheric aerosols (FIOCCO and VISCONTI, 1973) or meridional transport (FUJIWARA and HIRONO. 1973) are similarly difficult to test.

DYNAMICAL EFFECTS

A number of workers have pointed to the existence of perturbations in the sodium layer which appear to be caused by dynamical effects in the main atmospheric constituents. As has already been mentioned, CLEMESHA et al. (1982) have demonstrated the existence of strong diurnal and semidiurnal oscillations in the layer, attributable to tidal perturbations. Shorter period oscillations, apparently caused by gravity waves propagating through the layer, have been studied by RICHTER et al (1981). THOMAS et al. (1977) and CLEMESHA et al. (1981c) have made spaced observations with a steerable lidar in order to investigate horizontal gradients, and the latter workers have used such measure-
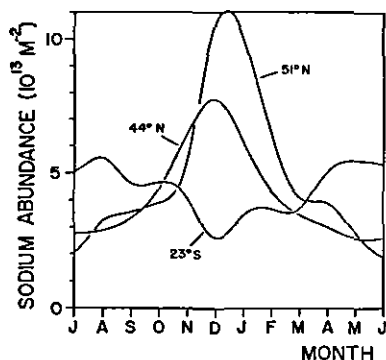

Figure 7. Seasonal variations of sodium for three latitudes.

ments to derive horizontal wind velocities in the 80 to 100 km region.

TEMPERATURE MEASUREMENTS

The ground state alkali metal atoms detected by laser radar are in thermal equilibrium with the ambient air molecules, so a measurement of their velocity distribution is equivalent to a determination of the ambient temperature. If, as is usually the case, the laser bandwidth is considerably greater than the width of the thermal Doppler spectrum (about 0.03 Å for sodium), then it would be possible, in principle, to determine the temperature by measuring the spectrum of the returned signal. In practice, a direct determination of this sort would be very difficult because of the very low light levels involved and the high spectral resolution necessary. BLAMONT et al. (1972) overcame this problem by measuring the attenuation produced in the returned signal by a sodium vapor cell maintained at a known temperature. Since the extinction depends on the linewidth of the signal returned from the sodium layer, its measurement can be used to derive the temperature of the sodium atoms. This technique was first used in sodium twilight measurements by BRICARD and KASTLER (1944).

The absorption cell technique used by BLAMONT and his coworkers suffers from the disadvantage that the extinction produced by the cell is a rather slowly varying function of the temperature of the scattering atoms responsible for the lidar signal, with the result that the latter must be integrated over long time periods in order to give the required precision of measurement. An alternative technique, which has been demonstrated by THOMAS and BHATTACHARYYA (1980), is to use a very narrow bandwidth laser to actually measure the thermal velocity distribution by measuring the signal return at a number of discrete wavelengths. This technique, requiring as it does, very precise control of the laser wavelength, involves certain experimental difficulties, but offers the potential for accurate determination of temperature in the 80 to 100 km range with good resolution in both height and time.

FUTURE WORK

As a result of lidar observations, it is probably true to say that we now know more about the morphology of atmospheric sodium than any other minor constituent of the earth's upper atmosphere. On the other hand, this knowledge has not led to a greatly increased understanding of the origin of atmospheric sodium, nor of the way in which it interacts with other species. The main difficulties in the way of solving these problems are a lack of accurate knowledge of the variations in space and time of other minor constituents, and of the appropriate rate coefficients for reactions between sodium and these species, combined with our ignorance of both vertical and horizontal transport effects. In view of this situation it seems that an improved understanding of the alkali metals in the atmosphere will come about not so much from better measurements of their variations in space and time, but rather from improved measurements of the other species with which they might react, better laboratory determinations of the appropriate rate coefficients, and an improved understanding of transport effects. This should not be taken to imply that more observations are unnecessary, particularly in the case of the alkali metals other than sodium, for which few measurements have been made. Improved geographical coverage would be helpful. All the observations so far having been made at middle to high latitudes in the northern hemisphere, or low latitude in the southern. The possible existence of an asymmetry in the seasonal variation between northern and southern hemisphere.

The proposed Shuttle Lidar (NASA, 1979) could make a major contribution to our understanding of the alkali metals in the atmosphere. Not only would the improved geographical coverage be important, but the proposed measurement of

other minor species such as O_3, and OH might greatly improve our understanding of this region of the atmosphere.

Apart from the potential of the Shuttle Lidar, it is the opinion of the author that the most important future contribution of lidar observations of the alkali metals lies in the use of these species, principally sodium, as tracers for atmospheric dynamics and temperature measurements. In the height region between 80 and 100 km, an adequately designed and located lidar could provide continuous measurements of temperature with a time resolution of about 1 km, along with estimates of horizontal and vertical velocity components. Such a system would be a most powerful tool for the study of atmospheric dynamics in the mesopause region.

ACKNOWLEDGEMENTS

REFERENCES

Bain, W. C. and M. C. W. Sandford (1966), Light scatter from a laser beam at heights above 40 km, J. Atmos. Terr. Phys., 28, 543-552.

Blamont, J. E., M. L. Chanin and G. Megie (1972), Vertical distribution and temperature profile of the nighttime atmospheric sodium layer obtained by laser backscatter, Ann. Geophys., 28, 833-838.

Blamont, J. E. and T. M. Donahue (1964), Sodium dayglow: observation and interpretation of a large diurnal variation, J. Geophys. Res., 69, 4093-4127.

Bowman, M. R., A. J. Gibson and M. C. W. Sandford (1969), Atmospheric sodium measured by a tuned laser radar. Nature, 221, 456-457.

Bricard, J. and A. Kastler (1944), Recherches sur la radiation D du sodium dans la lumiere du ciel crepusculaire et nocturne, Ann. Geophys., 1, 53.

Clemesha, B. R., G. S. Kent and R. W. Wright (1966), Laser probing the lower atmosphere, Nature, 209, 184.

Clemesha, B. R., D. M. Simonich and V. W. J. H. Kirchhoff (1975), Automatic wavelength control of a flashlamp pumped dye laser, Optical and Quantum Electronics, 7, 193.

Clemesha, B. R., V. W. J. H. Kirchhoff, D. M. Simonich, H. Takahashi and P. P. Batista (1980), Spaced lidar and nightglow observations of an atmospheric sodium enhancement. J. Geophys. Res., 85, 3480-3484.

Clemesha, B. R., V. W. J. H. Kirchhoff and D. M. Simonich (1981a), Remote measurements of tropospheric and stratospheric winds by ground-based lidar, Applied Optics, 20, 2907-2910.

Clemesha, B. R., V. W. J. H. Kirchhoff and D. M. Simonich (1981b), Comments on "A meteor ablation model of the sodium and potassium layers" by D. M. Hunten, Geophys. Res. Lett., 8, 1023-1025.

Clemesha, B. R., P. P. Batista, V. W. J. H. Kirchhoff and D. M. Simonich (1981c), Mesospheric winds from lidar observations of atmospheric sodium, J. Geophys. Res., 86, 868-870.

Clemesha, B. R., D. M. Simonich, P. P. Batista and V. W. J. H. Kirchhoff (1982), The diurnal variation of atmospheric sodium, J. Geophys. Res., 87, 181-186.

Collis, R. T. H. and M. G. H. Ligda (1966), Note on lidar observations of particulate material in the stratosphere, J. Atmos. Sci., 23, 255-257, 1966.

Elterman, L. (1954), Seasonal trends of temperature, density and pressure to 67.6 km obtained with the searchlight probing technique, J. Geophys. Res., 59, 351-358.

Felix, F., W. Keenliside, G. S. Kent and M. C. W. Sandford (1973), Laser radar observations of atmospheric potassium, Nature, 246, 345.

Fiocco, G. and L. D. Smullin (1963), Detection of scattering layers in the upper atmosphere (60-140 km) by optical radar, Nature, 199, 1275-1276.

Fiocco, G. and G. Visconti (1973), On the seasonal variation of upper atmospheric sodium, J. Atmos. Terr. Phys., 35, 165-171.

Friedland, S. S., J. Katzenstein and M. M. Zatzick (1956), Pulsed searchlighting the atmosphere, J. Geophys. Res., 61, 415.

Fujiwara, M. and M. Hirono (1973), A theory for seasonal variation of the meteoric element densities in the upper atmosphere (II), Report of Ionosphere and Space Research in Japan, 27, 223-227.

Gibson, A. J. and M. C. W. Sandford (1971), The seasonal variation of the nighttime sodium layer, J. Atmos. Terr. Phys., 33, 1675-1684.

Gibson, A. J. and M. C. W. Sandford (1972), Daytime laser radar measurements of the atmospheric sodium layer, Nature, 239, 509-511.

Granier, C. and G. Megie (1982), Daytime lidar measurements of the mesospheric sodium layer, Planet. Space Sci., 30, 169-177.

Hulburt, E. O. (1937), Observations of a searchlight beam to an altitude of 28 km, J. Opt. Soc. Amer., 27, 377-382.

Hunten, D. M. (1967), Spectroscopic studies of the twilight airglow, Space Sci. Rev., 6, 493-573.

Jegou, J., M. Chanin, G. Megie and J. E. Blamont (1980), Lidar measurements of atmospheric lithium, Geophys. Res. Lett., 7, 995-998.

Johnson, E. A., R. C. Meyer, R. E. Hopkins and W. H. Mock (1939), The measurement of light scattered by the upper atmosphere from a searchlight beam, J. Opt. Soc. Amer., 29, 512-517.

Kent, G. S., B. R. Clemesha and R. W. Wright (1967), High altitude atmospheric scattering of light from a laser beam, J. Atmos. Terr. Phys., 29, 169-181.

Kirchhoff, V. W. J. H. (1983), Atmospheric sodium chemistry and diurnal variations: an up-date, Geophys. Res. Lett., in press.

Kirchhoff, V. W. J. H., B. R. Clemesha and D. M. Simonich (1981), The atmospheric neutral sodium layer. 1, recent modeling compared to measurements, J. Geophys. Res., 86, 6892-6898.

Kirchhoff, V. W. J. H. and B. R. Clemesha (1983), The atmospheric neutral sodium layer, II, Diurnal variations, J. Geophys. Res., 88, 442-450.

Maiman, T. H. (1960), Stimulated optical radiation in ruby, Nature, 187, 493-494.

Megie, G. and J. E. Blamont (1977), Laser sounding of atmospheric sodium: interpretation in terms of global atmospheric parameters, Planet. Space Sci., 25, 1093-1109.

Megie, G., F. Bos, J. E. Blamont and M. L. Chanin (1978), Simultaneous nighttime lidar measurements of atmospheric sodium and potassium, Planet. Space Sci., 26, 27-35.

McCormick, P. D., S. K. Poultney, U. Von Wijik, C. O. Alley, R. T. Bettinger and J. A. Perschy (1966), Backscattering from the upper atmosphere (75-160 km) detected by optical radar, Nature, 209, 798-799.

NASA, Shuttle Atmospheric Lidar Research Program - Final Report of the Atmospheric Lidar Working Group, NASA SP-433, 1979.

Richter, E. S., J. R. Rowlett, C. S. Gardner and C. F. Sechrist (1981), Lidar observation of the mesospheric sodium layer over Urbana, Illinois, J. Atmos. Terr. Phys., 43, 327-337.

Simonich, D. M., B. R. Clemesha and V. W. J. H. Kirchhoff (1979), The mesospheric sodium layer at 23°S.: nocturnal and seasonal variations, J. Geophys. Res., 84, 1543-1550.

Simonich, D. M. and B. R. Clemesha (1983), Resonant extinction of lidar returns from the alkali metal layers in the upper atmosphere, Appl. Optics, 22, 1387-1389.

112

Synge, E. H. (1930), a method of investigating the higher atmosphere, Phil. Mag., 9, 1014.

Thomas, L., and S. K. Bhattacharyya (1980), Mesospheric temperatures deduced from laser observations of the sodium D_2 line profile, Proc. of the Vth ESA-PAC Symposium on European Rocket and Balloon Programmes and Related Research – Bournemouth (UK), June.

Thomas, L., A. J. Gibson and S. K. L. Bhattacharyya (1977), Lidar observations of a horizontal variation in the atmospheric sodium layer, J. Atmos. Terr. Phys., 39, 1405-1409.

9. PARTIAL REFLECTION D-REGION ELECTRON DENSITIES

A. H. Manson and C. E. Meek
Institute of Space and Atmospheric Studies
University of Saskatchewan
Saskatoon, Sask., Canada

INTRODUCTION

The most common name for this technique of measuring electron densities N as a function of height, h, in the D region (~60-90 km) is the differential absorption technique (DAE). GARDNER and PAWSEY (1953) were the first to make measurements of medium frequency ordinary (0) and extra-ordinary (X) waves reflected from the D region, in order to calculate electron densities. This original paper (hereafter G & P) is elegant in its physical and mathematical form and if used as the only source of information on the technique it would provide the experimenter of today with almost all the essential knowledge to obtain good N(h) profiles.

Groups of scientists in a number of countries have tested the many assumptions of G & P, and have sought to improve their experiment (e.g. by adding differential phase), and have made it much more complicated (volume scatter theories; deconvolution)! The main scientific groups involved in the 20 years from ~1953-1973 were from Canada (BELROSE and BURKE, 1964; BELROSE, 1970 (a review)), Norway (THRANE et al. 1968), New Zealand (GREGORY and MANSON, 1968, 1969a), and U.S.A. (FLOOD, 1968). In the last ten years there has been less attention to the proving of the experiment and more to its use for ionospheric and dynamical studies. The most active groups include the University of Canterbury, N.Z. (H. A. VON BIEL, G. J. FRASER); the University of Adelaide, Australia (R. A. VINCENT); the University of Illinois, U.S.A. (S. A. BOWHILL); the University of Sasketchewan, Canada (C. E. MEEK, A. H. MANSON), the University of Tromso, Norway (A. BREKKE, A. HAUG, O. HOLT, E. V. THRANE); also until the mid 1970's J. S. BELROSE and colleagues from CRC, Ottawa, Canada; and most recently J. PRIESE, W. SINGER of the Central Institute of Solar-Terrestrial Physics (Juliusruh, GDR, 55°N, 19°E).

Papers from these various groups will be discussed below: only highlights can be mentioned, but hopefully all major contributors to this technique will be mentioned (at least once). Private communications with present key workers (including J. S. BELROSE, CRC, Ottawa, Canada) is highly recommended.

PRINCIPLES OF THE TECHNIQUE

In the basic experiment, pulses of medium or high frequency, usually at a fixed frequency (2-6 MHz), are radiated upwards with known wave polarization (usually linear or circular) from a transmitter at ground level. Typical system parameters are shown in Table 1.

Table 1

Transmitter:	50-100 kW peak power, ~25 μs pulse width 6-60 Hz repetition rate.
Receiver:	60 kHz bandwidth, 2-5 dB noise factor.
Transmitting	Receiving Array: Dipoles arranged orthogonally ~$\lambda/4$ above the ground (2 x 2 arrays have been used but larger arrays are preferred to narrow the beam).

Partial reflections, from ionospheric "scatterers" at heights below the E region (~60-90 km), are received at the ground, and are resolved into two characteristic components, the ordinary (O) and extraordinary (E) modes whose amplitude ratio A_x/A_o is then measured as a function of height, h. The heights of these scatterers are determined by delay times, the group retardation being minimal in the undisturbed D-region. The electronic system can be very simple. Power splitters and quadrature networks to separate the A_x and A_o components are commercially available at low prices and an A-D converter, height-gate system, and micro-computer allows the real-time calculation of mean amplitudes. Typical gate spacings would be 2-3 km.

It is remarkable that the nature and cause of the radiowave scatterers is still not fully understood. Indeed a separate paper could be written on that topic alone. GARDNER and PAWSEY assumed, after discussion, that they were sharply bounded horizontally stratified irregularities in N(h), of Fresnel zone scale. This seemingly implausible notion has been explored for 30 years and especially in the last 10 years (GREGORY, 1956, 1961; PIGGOTT and THRANE, 1966; MANSON, 1966; HOLT, 1969; FLOOD, 1969; MANSON et al. 1969; GREGORY and VINCENT, 1970; VINCENT, 1973; VINCENT and BELROSE, 1978; HILL and BOWHILL, 1979; HOCKING, 1979; THRANE et al., 1981; HOCKING and VINCENT, 1982). The frequent conclusion is that the assumption is a physically realistic one below 80 km (depending on latitude and season) and that even above 80 km, where more isotropic turbulent processes become important, it is as useful as the more complex "volume scatter" assumptions - at least within the theoretical development required to calculate N(h).

The ratio of the coefficients of reflection, R_x/R_o, of the two modes, as they originate at each reflection height is then calculable, using the SEN-WYLLER (1960) magneto-ionic equations (a significant change from the non-generalized equations of G & P). Provided that the refractive index, n, for the radiowave is close to unity, then R_x/R_o is independent of N(h); departures typically become important for heights ≥ 85 km and N(h) of $\geq 4 \times 10^3$ cm^{-3} The ratio of A_x/A_o is related to R_x/R_o by

$$A_x/A_o = R_x/R_o \, \exp \{-2 \int_o^h (k_x - k_o) \, dh\} \tag{1}$$

so that

$$N(h) = \frac{\frac{d}{dh} (\ln \frac{R_x}{R_o}) - \frac{d}{dh} (\ln \frac{A_x}{A_o})}{\frac{2 (k_x - k_o)}{N}} \tag{2}$$

where $(k_x - k_o)$ is the normalized differential absorption $(\chi_x - \chi_o) \, \omega/c$ and ω is the wave frequency, and $\chi_{x,o}$ the imaginary part of the complex refractive index $(\mu - i\chi)$. The normalized differential absorption may also be shown to be independent of N(h) for similar h and N(h).

A similar equation may be used if the phase ϕ of the wave is measured as a function of height so that:

$$N(h) = \frac{\frac{d}{dh} (\phi_{R_x} - \phi_{R_o}) - \frac{d}{dh} (\phi_x - \phi_o)}{\frac{2\omega}{c} \frac{\mu_x - \mu_o}{N}} \tag{3}$$

This method is known as the differential phase experiments (DPE).

There has been considerable discussion of how to best form the A_x/A_o curve as a function of height, and to find the slope of that line. Mean amplitudes (from root mean square calculations) rather than mean $\overline{A_x}/A_o$ ratios are preferred (BELROSE, 1970), and care is required in finding slopes at certain heights and under certain disturbed conditions (BELROSE, 1970; GREGORY and MANSON, 1973; HARGREAVES and BREKKE, 1981; MEEK, 1977; MEEK and MANSON, 1981).

It is noted that a collision frequency profile $\nu(h)$, is required for equations (1) and (2). This is either obtained from atmospheric pressure (p) profiles (e.g. GREGORY and MANSON, 1969a)

$$\nu(h) = (6.4 \pm 0.4) \times 10^7 \, p \, \sec^{-1} \tag{4}$$

where p is in mb. Similar relations are given by other workers. A value for ν may be obtained from equation (1) when $R_x/R_o = A_x/A_o$ at low heights, alternatively (3) and (4) may be solved simultaneously to obtain $\nu(h)$. The latter is not a simple procedure overall. Fortunately, the calculated $N(h)$ is not strongly dependent upon $\nu(h)$, so this is one of the smaller sources of error (GREGORY and MANSON, 1969a). However, seasonal studies of $N(h)$ should include seasonally varying $\nu(h)$ profiles (GREGORY and MANSON, 1969b), obtained from standard atmospheres.

DISCUSSION OF THE TECHNIQUE

It is important to summarize the various tests of the basic theory including alternate theoretical approaches and developments which are in use.

IRREGULARITY STRUCTURE

The possibility of the irregularities having a perturbation in $\nu(h)$ as well as or instead of $N(h)$ were considered theoretically by PIGGOTT and THRANE (1966) and MANSON (1966). BELROSE et al. (1972) and MEEK and MANSON (1981) measured the ratios (α) of collision frequency to electron density fractional changes. Both theory and observations generally favor the dominance of $N(h)$ perturbations, and all groups now assume that.

VOLUME SCATTER

A major controversy began in the 1960's with the work of FLOOD (1968) and VON BIEL et al. (1970) who developed a theory based on "volume scattering", in which reflection and differential absorption coefficients were necessarily required throughout the pulse width. The method requires iteration from the base of the ionosphere and is sensitive to errors in measurement. Indeed although the work of HOCKING and VINCENT (1982) favors smaller more isotropic scatterers above ~80 km, it is not clear that all FLOOD's assumptions are fully justified. HOLT (1969) argued against the necessity for such complexity of method even if volume scattering occurs, and showed that providing heights were measured to the center of returned pulses, the simple G & P theory could be applied with adequate accuracy. (See FLOOD (1969) for a reply). BELROSE (1970) and AUSTIN and MANSON (1969) considered this and other evidence from variable pulse width experiments, to reach a similar conclusion, i.e. that the discrete Fresnel discontinuity assumption is adequate.

Another approach was taken by COYNE and BELROSE (1973) in which they assumed statistically independent scatterers within the sampled volume, and which used a deconvolution algorithm to find the "real" A_o, A_x profiles. The is not suitable for routine processing (I. A. BOURNE, private communication). MONTBRIAND and BELROSE (1976) have discussed selection criteria whence a smooth

curve may be drawn through the "selected correct ratios" of A_x/A_o, occurring at echo peaks or where amplitudes vary slowly with height.

HEIGHTS OF APPLICABILITY ($\mu \approx 1$)

As already noted the basic G & P theory is appropriate to conditions where $\mu \approx 1$, or N(h) is less than a few thousand electrons per cm^3. MEEK (1977) considered differential pulse delay between O and X modes (as did DA SILVA and BOWHILL, 1974) and also horizontal displacement of the reflection points. These various effects all become serious for N(h) $\geq 4 \times 10^3$ cm^{-3} at 2.2 MHz – the heights would vary from ~90 km for quiet and <80 km for disturbed conditions. In Figures 1 and 2 we show the N(h), v(h) profiles used by MEEK and MANSON (1981) and the function $(k_x - k_o)/N$ from Equation 2 for various values of N(h).

This variation raises a very serious difficulty in interpreting A_x/A_o curves in ionospherically disturbed conditions: because of auroral disturbances this problem has been addressed in some detail by Canadian workers (MONTBRIAND and BELROSE, 1976; KHAN and MANSON, 1974; MANSON and KHAN, 1974; CHAKRABARTY et al., 1983) and Norwegian workers (HUAG et al., 1977; HARGREAVES and BREKKE, 1981; RASTOGI et al., 1982). The morphology of the precipitation event must be followed – weakening E-region signals, decreasing A_x/A_o value and loss of D-region signals – and other instruments, such as riometers, ionosondes and magnetometers should be used.

OBLIQUE REFLECTIONS

A major problem, whose severity varies with geographical location, height and time of day, is the presence of oblique echoes. The transmitter antenna at Saskatoon has a half beamwidth of ~22°, typical of many other systems, and receipt of echoes from actual heights below the nominal can be serious. In Figure 3 (MEEK, 1977) A_x/A_o is shown to be almost constant from 68–80 km at noon, and unrealistic values of electron density ~30 cm^{-3} are calculated.
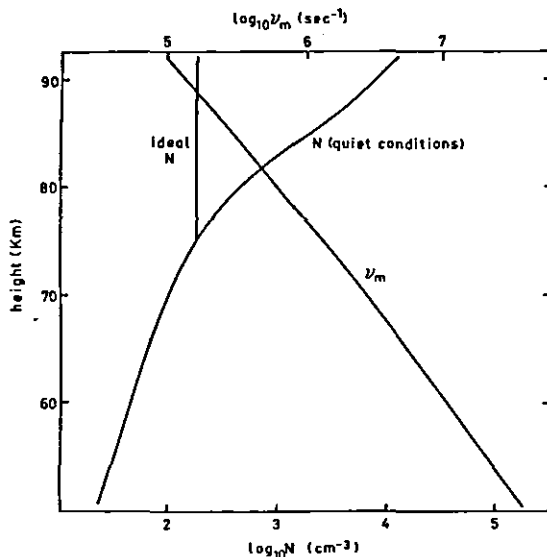

Figure 1. The ionospheric model used in
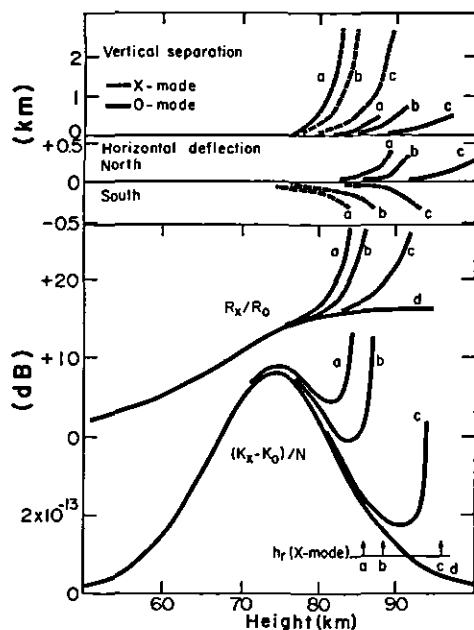MEEK and MANSON (1981).

Figure 2. The normalized differential absorption, also vertical and horizontal displacements for O and X modes. The $N(h)$, $\nu_m(h)$ used are from Figure 1 (and also $N(h) \times 5$, $\times 10$).
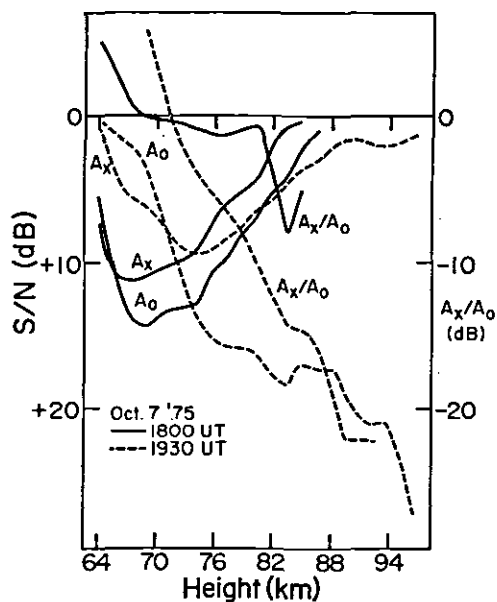

Figure 3. The measured A_X, A_O, A_X/A_O profiles for two observing intervals at Saskatoon.

Later at 1:30h (LT) a more normal situation applies with N(h) ~ 300 cm^{-3}. Au-
roral electron precipitation probably occurred at noon. Similar effects are
reported from the 2.75 MHz radar near Tromso, Norway (references above) although
they do not stress the obliquity problem.

MEEK (1977) studied all aspects of the problems associated with obliquity
such as the projection of ellipses onto the horizontal as a function of azimuth
angle, and magnetic field direction, and the actual lower heights of oblique
echoes having identical ranges. The latter effect is the dominant one. We
show in Figure 4 (MEEK and MANSON, 1981) calculations of A_x/A_o and $(\phi_x - \phi_o)$
various situations: "vertical" is the ideal profile, "measured" is for the
realistic case of equal scatter from 0-25°. The effect can clearly be serious,
especially as scatter is often found to be due to turbulent as well as specular
processes above ~80 km (HOCKING and VINCENT, 1982). THRANE et al. (1981) found
echoes considered to be of an isotropic character (due to turbulence) at all
D-region heights in Norway.

This is one reason why final real-time calculations of N(h) is seldom com-
pleted at observatories; study of A_x, A_o and A_x/A_o profiles, diurnal
of observations and the use of magnetometers, riometers and solar/particle
satellite data are often needed before final selection of N(h) values can be
made (e.g. CHAKRABARTY et al., 1982, 1983).

NOISE CORRECTION/OVERLOADING

Considerable care is required here, especially as the $A_o \gg A_x$ at high
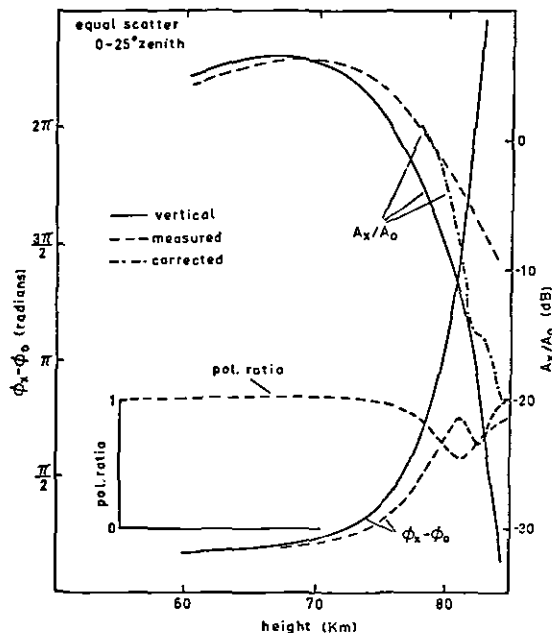

Figure 4. A calculation of the effect of
oblique scatter on the "measured" DAE,
DPE parameters. Polarization ratios for
an FPE are also shown. The figure is
explained fully in the text.

altitudes: BELROSE (1970), VON BIEL (1971, 1979), DA SILVA and BOWHILL (1974),
MEEK (1977).

NEW DEVELOPMENTS - POLARIMETERS

    Two groups are using polarimeters, which provide the linear, diagonal and
circular waves. This enables a complete description of the wave: major and
minor axes of the ellipse, and tilt angle (related to Faraday rotation). From
these, A_x/A_o and $\phi_x - \phi_o$ are available. Another description is by the
parameters (BORN and WOLF, 1965, p. 555).

    VON BIEL (1977) has used a polarimeter to provide more accurate determina-
tions of the A_x/A_o ratios: his method, which also includes volume scatter-
ing, is elegant if somewhat complex. The polarimeter used by MEEK and MANSON
(1981) is simple to develop from a basic DAE system, by adding a commercial
power splitter and hybrid-junction network (Figure 5). The theory is outlined
in their paper. They used a polarimeter to measure the unpolarized power:
this can be due to external noise and independent scatterers in the volume which
contribute different A_x/A_o and $\phi_x - \phi_o$ values during measurement of mean square
amplitudes for an N(h) profile. Volume scatter of obliques could be responsi-
ble. The theoretical curves of Figure 4 show that the unpolarized component
increases (P < 1) near 80 km where the "measured" A_x/A_o departs significant
from the ideal curve. "Corrected" ratios (obtained by subtracting the unpolar-
ized noise power from A_x, A_o) can lead to overcorrection in the slope of A_x/A_o
for this case of severe oblique contamination. Actual data show a similar ef-
fect. We consider the full polarization experiment (FPE) to be most valuable
in identifying, rather than in correcting for, conditions under which the normal
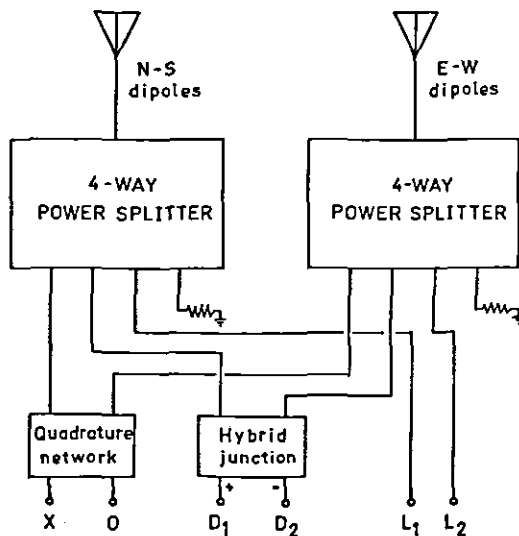DAE theory cannot be used to determine N(h), because of oblique reflections or
volume scatter.

Figure 5. The experimental arrangement
used to create the six required compo-
nents of the received wave from two
sets of orthogonal dipoles.

NEW DEVELOPMENT - CHIRP RADARS

It is very difficult to use MF radars in Europe, due to noise and other "short-wave" transmitters, but also and most seriously because licenses for such high power narrow pulse devices will not be issued. PREISE and SINGER (MAP Symposium - Schwerin GDR, May 1983) report the successful use of the FM-CW-radar principle at Juliusruh (55°N, 13°E) near the Observatory of Ionospheric Research, Kuhlungsborn (GDR). The 1 kW linear FM-chirp of 325 kHz bandwidth centered on 3.18 MHz is Gaussian weighted in amplitude and gives a height resolution of 1.5 km. Excellent N(h) profiles are being obtained from 70-90 km, and wind measurements are planned. This development allows for the possibility of the use of MF radar in locations of Europe and elsewhere which were unavailable before.

CONCLUSIONS

The use of MF radars to obtain N(h) from ~60-90 km, using DAE, DPE or FPE methods is an extremely valuable and economical technique. It is not an easy technique to use however, and interpretation depends critically on the nature of the scattering layers/regions and ionospheric disturbances. Height resolution may be as good as 2-3 km, but accuracies may vary from a few percent (e.g. MONTBRIAND and BELROSE, 1972 - during an X-ray solar flare) to >100% if conditions are disturbed or obliques/volume scatter are undiagnosed. A good argument can be made (I. A. BOURNE, A. H. MANSON private communication) that the best MF system would involve high power, narrow pulses and narrow antenna beamwidths: however such a design is environmentally noisy and expensive. For most existing systems, analysis is never routine!

A great deal of excellent ionospheric physics has resulted from the use of the partial reflection systems however (the list below is not exhaustive and is meant only to show the range of explored phenomena):

1. Seasonal variations in N(h) (winter anomaly): GREGORY and MANSON (1969b), COYNE and BELROSE (1972), CHAKRABARTY et al. (1983).

2. Variations in N(h) due to dynamics/waves/winds: GREGORY and MANSON (1970), WRATT (1977), WEILAND and BOWHILL (1978, 1981), MEEK and MANSON (1978).

3. Seasonal variations in electron loss coefficient ($\psi = q/N^2$): MONTBRIAND and BELROSE (1979).

4. Electron precipitation (effects and derived spectra): KHAN and MANSON (1974), MANSON and KHAN (1974), LARSEN et al. (1976a,b), MONTBRIAND and BELROSE (1976), RASTOGI et al. (1982), CHAKRABARTY et al. (1983).

5. Comparison with photochemical models (diagnostic): Chakrabarty et al. (1982).

The advantage of the MF radars, is that in addition they may be used to study scatter processes/turbulence (e.g. THRANE et al. (1981), MANSON et al. (1981), HOCKING and VINCENT, (1982), HOCKING, (1983a,b), and by means of spaced antennas or doppler measurements (narrow beam antennas), winds and wave motions (60-110 km). These latter are discussed elsewhere in this volume.

REFERENCES

Austin. G. L. and A. H. Manson (1969), On the nature of irregularities that produce partial reflections of radio waves from the lower ionosphere (70-100 km). Radio Sci., 4, 35-40.
Belrose. J. S. (1970), Radio wave probing of the ionosphere by the partial re-

flection of radio waves (from heights below 100 km), J. Atmos. Terr. Phys., 32, 567-596.

Belrose, J. S. and M. J. Burke (1964), Study of the lower ionosphere using partial reflection, I. Experimental technique and methods of analysis, J. Geophys. Res., 69, 2799-2818.

Belrose, J. S., M. J. Burke, T. N. R. Coyne and J. E. Reed (1972), D-region measurements with the differential-absorption, differential-phase partial-reflection experiments, J. Geophys. Res., 77, 4829-4838.

Born. M. and E. Wolfe (1965), "Principles of Optics", 3rd edition (revised), Pergamon, Oxford.

Chakrabarty. P., C. E. Meek, D. K. Chakrabarty and A. H. Manson (1982), Results inferred from electron density measurements at Saskatoon, Canada (L = 4.4) by a partial reflection technique - I. Variations of nitric oxide in the D-region during quiet periods, J. Atmos. Terr. Phys., 44, 957-962.

Chakrabarty. P., C. E. Meek, D. K. Chakrabarty and A. H. Manson (1983), Results inferred from electron density measurements at Saskatoon, Canada (L=4.4) by a partial reflection technique - II. Ion production rates and nitric oxide in the D-region during post-storm periods, J. Atmos. Terr. Phys., 45, 303.

Coyne. T. N. R. and J. S. Belrose (1972), The diurnal and seasonal variation of electron densities in the mid-latitude D-region under quiet conditions, Radio Sci., 7, 163-174.

Coyne. T. N. R. and J. S. Belrose (1973), An investigation into the effects of limited height resolution in the differential-absorption partial-reflection experiment, J. Geophys. Res., 78, 8276-8288.

Da Silva. L. S. and S. A. Bowhill (1974), Aeronomy Rept. No. 63, Aeronomy Lab., Dept. of Elec. Eng., Univ. of Illinois.

Flood, W. A. (1968), Revised theory for partial reflection D-region measurements, J. Geophys. Res., 73, 5585-5598.

Flood, W. A. (1969), Reply. J. Geophys. Res., 74, 5183-5186.

Gardner, F. F. and J. L. Pawsey (1953), Study of the ionospheric D-region using partial reflections, J. Atmos. Terr. Phys., 3, 321-344.

Gregory. J. B. (1956), Ionospheric reflections from heights below the E-region, Aust. J. Phys., 9, 324-342.

Gregory. J. B. (1961), Radio wave reflections from the mesosphere, I. Heights of occurrence. J. Geophys. Res., 66, 429-445.

Gregory. J. B. and A. H. Manson (1968), Discussion of letter by A. J. Ferraro and H. S. Lee, 'Capability of a High-Power wave interaction facility'. J. Geophys. Res., 66, 429-445.

Gregory. J. B. and A. H. Manson (1969a), Seasonal variations of electron densities below 100 km at mid-latitudes - I. Differential absorption measurements, J. Atmos. Terr. Phys., 31, 683-701.

Gregory, J. B. and A. H. Manson (1969b), Seasonal variations at electron densities below 100 km at mid-latitudes - II. Electron densities and atmospheric circulation, J. Atmos. Terr. Phys., 31, 703-729.

Gregory. J. B. and A. H. Manson (1970), Seasonal variations of electron densities below 100 km at mid-latitudes - III. Stratospheric-ionospheric coupling, J. Atmos. Terr. Phys., 32, 837-852.

Gregory. J. B. and A. H. Manson (1973), Further comments on a paper by David B. Newman. Jr., and A. J. Ferraro, 'Sensitivity study of the partial reflection experiment', J. Geophys. Res., 78, 7572-7575.

Gregory, J. B. and R. A. Vincent (1970), Structure of partially reflecting regions in the lower ionosphere, J. Geophys. Res., 75, 6387-6389.

Hargreaves. J. K., A. Brekke, E. V. Thrane and K. Bjorna (1981), Application of the Ramfjordmoen partial reflection experiment to the study of auroral precipitation events, J. Atmos. Terr. Phys., 43, 1093-1106.

Haug, A., A. Brekke and O. Holt (1977), Observations of unusually strong partial reflections in the auroral D-region during an absorption event, J. Atmos. Terr. Phys., 39, 1333-1340.

Hill, R. J. and S. A. Bowhill (1979), Relaxation to photochemical equilibrium

of charged species within displaced air parcels in the D-region, J. Atmos. Terr. Phys., 41, 607-623.

Hocking, W. K. (1979), Angular and temporal characteristics of partial reflections from the D-region of the ionosphere, J. Geophys. Res., 84, 845-851.

Hocking, W. K. (1983a), On the extraction of atmospheric turbulence parameters from radar backscatter Doppler spectra - I. Theory, J. Atmos. Terr. Phys., 45, 89-102.

Hocking, W. K. (1983b), Mesospheric turbulence intensities measured with a HF radar at 35°S - II. J. Atmos. Terr. Phys., 45, 103-114.

Hocking, W. K. and R. A. Vincent (1982), Comparative observations of D Region HF partial reflections at 2 and 6 MHz, J. Geophys. Res., 87, 7615.

Holt, O. (1969), Discussion of paper by W. A. Flood, 'Revised theory for partial reflection D-region measurements', J. Geophys. Res., 74, 5179-5182.

Khan. Z. M. and A. H. Manson (1974), Electron precipitation at an auroral latitude-Saskatoon, L=4.4. I. Photometer, magnetometer and radiowave data, Can. J. Phys., 52, 1872-1878.

Larsen. T. R., W. L. Imhof and J. B. Reagan (1976b), L-dependent energetic electron precipitation and mid-latitude D region pair production profiles, J. Geophys. Res., 81, 2200-2212.

Larsen, T. R., W. L. Imhof and J. B. Reagan (1976b), L-dependent energetic electron precipitation and mid-latitude D region pair production profiles, J. Geophys. Res., 81, 3444-3446.

Manson, A. H. (1966), Comments of the theory of the differential absorption experiment, J. Geophys. Res., 71, 3783-3785.

Manson, A. H. and Z. M. Khan (1974), Electron Precipitation at an auroral latitude - Saskatoon, L=4.4 II. Fluxes of precipitating electrons, Can. J. Phys., 52, 1879-1884.

Manson, A. H., C. E. Meek and J. B. Gregory (1981), Gravity waves of short period (5-90 min), in the lower thermosphere at 52°N (Saskatoon, Canada); 1978/1979, J. Atmos. Terr. Phys., 43, 35-44.

Manson, A. H., M. W. J. Merry and R. A. Vincent (1969), Relationship between the partial reflection of radio waves from the lower ionosphere and irregularities as measured by rocket probes, Radio Sci., 4, 955-958.

Meek, C. E. (1977), Ph.D. Thesis, Physics Dept., University of Saskatchewan, Canada.

Meek, C. E. and A. H. Manson (1978), Comparisons between time variations in D-region winds and electron densities at Saskatoon, Canada (52°N; 106°W), J. Atmos. Terr. Phys., 40, 1267-1274.

Meek, C. E. and A. H. Manson (1981), Use of the full polarization measurement in the partial reflection experiment, J. Atmos. Terr. Phys., 43, 45-58.

Montbriand, L. E. and J. S. Belrose (1972), Effective electron loss rates in the lower D-region during the decay phases of solar X-ray (SxR) flare events, Radio Sci., 7, 133-142.

Montbriand, L. E. and J. S. Belrose (1976), Changes in electron precipitation inferred from spectra deduced from D-region electron densities during a post-magnetic storm effect, J. Geophys. Res., 81, 2213-2222.

Montbriand, L. E. and J. S. Belrose (1979), Diurnal and seasonal variations of the steady state loss coefficient in the D region, J. Geophys. Res., 84, 1921-1929.

Piggott, W. R. and E. V. Thrane (1966), The effect of irregularities in collision frequency on the amplitude of weak partial reflections, J. Atmos. Terr. Phys., 28, 311-314.

Rastogi. P. K., A. Brekke. O. Holt and T. Hansen (1982), Variability of D-region electron densities at Tromso, J. Atmos. Terr. Phys., 44, 313-323.

Thrane. E. V., B. Grandal, T. Fla and A. Brekke (1981), Fine structure in the ionospheric D-region, Nature, 292, 221-223.

Thrane, E. V., A. Haug, B. Bjelland, M. Anastassiades, and E. Tsagakis (1968), Measurements of D-region electron densities during the International Quiet Sun Years, J. Atmos. Terr. Phys., 30, 135-150.

Vincent, R. A. (1973), The interpretation of some observations of radio waves scattered from the lower ionosphere, Aust. J. Phys., 26, 815–827.

Vincent, R. A. and J. S. Belrose (1978), The angular distribution of radio waves partially reflected from the lower ionosphere, J. Atmos. Terr. Phys., 40, 35–47.

Von Biel. A. H. (1971), Amplitude distributions of D-region partial reflections, J. Geophys. Res., 76, 8365–8367.

Von Biel. A. H. (1977), The phase switched correlation polarimeter – a new approach to the partial reflection experiment, J. Atmos. Terr. Phys., 39, 769–778.

Von Biel. A. H. (1979), The use of amplitude distributions in practical data assessment problems, J. Atmos. Terr. Phys., 41, 1201–1204.

Von Biel. A. H., W. A. Flood and H. G. Camnitz (1970), Differential-phase partial-reflection technique for the determination of D-region ionization, J. Geophys. Res., 75, 4863–4870.

Weiland, R. M. and S. A. Bowhill (1978), Aeronomy Rept. No. 81. Aeronomy Lab., Dept. of Elec. Eng., University of Illinois.

Weiland, R. M. and S. A. Bowhill (1981), Aeronomy Rept. No. 100, Aeronomy Lab., Dept. of Elec. Eng., University of Illinois.

Wratt, D. S. (1977), Variations in electron density in the middle latitude D-region above Urbana, Illinois, J. Atmos. Terr. Phys., 39, 607–617.

10. MWR - METEOR WIND RADARS

R. G. Roper

School of Geophysical Sciences
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

The requirements of a "state of the art" meteor wind radar, and acceptable compromises in the interests of economy, are detailed. Design consideration of some existing and proposed radars are discussed. The need for international cooperation in mesopause level wind measurement, such as that being fostered by the MAP GLOBMET (Global Meteor Observations System) project, is emphasized.

INTRODUCTION

This is my second attempt to write a review of meteor wind radars. The first draft (in note form only) was very interesting, including as it did a history of meteor radar studies dating from the 1930's, and descriptions of the evolution of such radars into the sophisticated systems in use today. It also treated the theory of train formation and echo characteristics in some detail, together with an overview of the wind results produced, as they pertained to prevailing winds, tides, gravity waves and turbulence. To do justice to this outline would require several hundred pages, and would be, in fact, an update of D.W.R. McKinley's "Meteor Science and Engineering", published by McGraw Hill in 1961.

Here, I will concentrate on a description of the technique, with applications and limitations, and will provide descriptions of some current and proposed radars, with a bibliography which will provide additional information for those readers with specific interests.

THE TRACER

The upper atmosphere of the earth is continually being bombarded by meteoroids which, on encountering the increasing density of the atmosphere, incandesce and ablate, producing, if large enough, visible light and a trail of ionization. The relative "if large enough" can be quantified, at least conceptually, by saying "if larger than a grain of sand". It is the high velocity (between 11 and 72 kilometers per second geocentric, with an average value of 42 km/sec) of these small particles which makes possible their high ionization efficiency.

Because of the low atmospheric density, diffusion rates are high, and the process of ambipolar diffusion results in the average echo duration being less than a second at the VHF radar frequencies typically used. Lower radar transmitting frequencies will result in longer duration echoes (and will also raise the useable echo height ceiling), but have a greater potential for interference, and allocations with sufficient bandwidth for accurate range determination are difficult to obtain. Useable echo rate decreases as the transmitted frequency is raised; best frequencies for use are in the range from 30 to 40 MHz.

In general, individual trails will be less than 10 km long (of importance if a spaced station determination of wind shears is contemplated).

A few echoes per hour will be "overdense" (reflect incident radio wave as if the trail were a solid cylinder), and will produce more than one reflecting center as the trail is deformed by wind shears (radio reflection from a meteor

trail is specular, with the majority of the reflected energy coming from the Fresnel zone about the specular reflection point). Thus one should not expect to make reliable wind measurements at times beyond a second or two from initial trail detection. Trails which produce echoes rising slowly out of the noise are suspect — they are probably long duration trails assuming aspect sensitivity after formation.

The sharp rise in initial echo amplitude is a function of the meteor velocity (not the velocity of drift after trail formation, which is the line of sight component of the wind), and is accompanied by phase changes associated with the formation of successive Fresnel zones of the echo diffraction pattern. If phase measurements are attempted too early (particularly during the formation of the Fresnel zone around the specular reflection point), anomalous "wind" velocities will result. As a rule of thumb, wind doppler determinations should commence 40 to 50 milliseconds after the specular reflection point is passed by the meteoroid, and should not last more than 1 second (most echoes will decay into the noise in less than a second, anyway).

There is a diurnal variation in echo rate which is a consequence of the rotation of the earth as it proceeds in its orbit around the sun. At dawn, the maximum number of echoes occur because the observer is facing in the direction of travel of the earth, which "sweeps up" more debris than enters on the dusk side. This diurnal variation in echo rate can exceed 10:1 for systems using beam antennas; it will be less for "all sky" systems.

The height distribution of echoes detected by most systems is such that winds can be reliably determined over the height range 80 to 100 km.

The radio meteor trail is an ionized column drifting under the action of the neutral wind. Does this column drift as the neutral wind, or does the presence of the earth's magnetic field modify the motion? While theoretical discussion of this possibility continues, no evidence has yet been forthcoming that other than the neutral wind is being measured over the height range 80 to 100 km. Certainly, two distinct circulation patterns are seen, with, at mid latitudes (where most of the data to date has been collected), a "breakpoint" at 85 kilometers. However, the circulation above 85 kilometers is typical of the neutral thermosphere, with that below typical of the mesosphere.

MEASUREMENT PARAMETERS

In order to determine the variation with height and time of the wind field in the meteor region, the following parameters of each echo must be determined:

1. The line of sight drift of the trail;
2. the range of the echo;
3. the echo arrival angle;
4. the time of occurence of the echo.

From 2 and 3, the echo height can be determined. For prevailing winds and tides, accuracy should be within ±2 km.

Because of the short duration of the typical meteor echo, there is insufficient time available for sampling more than a fraction of a cycle of doppler for line of sight drifts less than 5 m/sec or so. Thus the better radars use a coherent but offset frequency (of order transmitted $f \pm 40$ Hz) to determine the line of sight drift. Such a procedure has an additional advantage in that the sense of drift (toward or away from the observer) is determined automatically. For prevailing winds or tides, accuracy of individual line of sight velocity determinations should be of the order of 2 to 3 m/sec.

CW OR PULSE

Both continuous wave and pulse techniques have been used to determine meso-
pause level winds from radio meteor trail drifts. Since the useable echo rate
is determined by average power, the CW technique has an advantage in that a
relatively simple transmitter with an average output of 300 watts (or more) can
be constructed and operated more economically than a pulsed radar having the
same average power output. Inexpensive receivers with low noise figure and
100 Hz bandwidth can be easily constructed, or bought "off the shelf". Range
can be measured by phase modulation of the transmitter, or by transmitting two
(or three) frequencies simultaneously, and comparing the phases of the dopplers
produced by each echo. The original Georgia Tech Radio Meteor Wind Facility
(ROPER, 1975) used two frequencies, 32.5 MHz ± 360 MHz, to determine range to
±2 km. The French CNET radars (see Appendix I) used a third frequency as a
vernier to determine range to better than 1 km.

The major disadvantages of the CW method are, (1) the necessity for
separation of transmitting and receiving sites, and (2) interference from air-
craft echoes. For fixed installations where air traffic is not a problem, a CW
system represents good value for money spent; the logistics of locating a
portable system are such that a single site pulsed radar will usually be
preferred, and will allow greater flexibility in siting (near airstrips, for
example). As regards interference created by the meteor radar, the potential
exists for either system, but is usually more easily cured in the CW case.

ANTENNA ARRAYS

The simplest system for the determination of the north/south (meridional)
and east/west (zonal) wind components consists of two yagi arrays looking at
right angles to each other, and at an elevation of 45°. Since reflection from
meteor trails is specular, and the majority of meteor orbits lie in the plane of
the ecliptic, it is better to look NE and NW (SE and SW in the souther hemi-
sphere) than due north and east. In order to maximize returns, instead of
looking for a predetermined time in each direction to determine hourly mean
winds, if a pulsed radar is used, each array can be alternately pulsed, with all
pulses going to one array when an echo has been sensed. Such a system is in use
at the Observatory of Ionospheric Research, Kuhlungsborn, GDR, where a "patrol"
pulse repetition refrequency of 25 Hz is used, increasing to 500 Hz when an echo
is detected. This has two advantages – average transmitter power is much lower
in the patrol mode, and such a low prf produces a long unambiguous range, which
is particularly necessary during solar sunspot maxima, when F region backscatter
can be bad at low VHF.

Such a simple system is adequate for the delineation of prevailing winds
and tides at the average echo height (∿95 km).

Various systems have been devised to measure echo arrival angle, enabling
the height of individual echoes to be determined. Some rely on relative
amplitude measurements (e.g., NOWAK et al., 1970), others on relative phase
(paper 5, Appendix I) and others on interferometry (RUDMAN et al., 1970; TSUDA,
1982).

A direction finding system devised by the Adelaide group (ROBERTSON et al.,
1953) illustrates the phase comparison technique. Although described here as a
CW application, the system may be adapted for use as a pulsed radar.

Antennae 1, 2 and 3 (see Figure 1) form an azimuth/elevation determining
set; they are aligned such that each receives the groundwave in phase (in this
scheme – all phases are established at the antennae, which means that phase
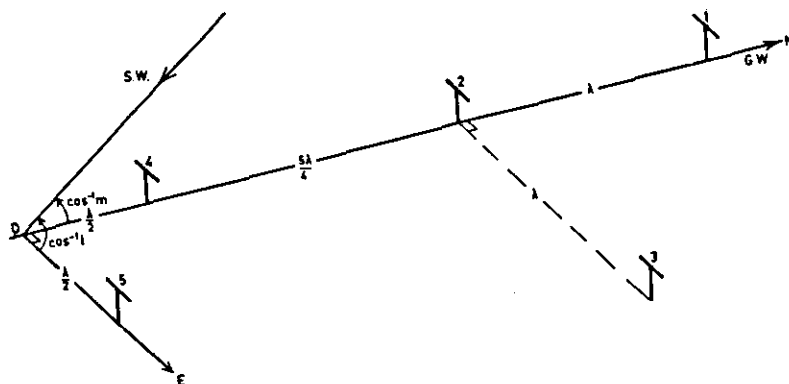coherence throughout the rest of the receiving system is relatively

Figure 1. Physical layout of the main receiving
station antenna array.

unimportant). Antennae 4 and 5 are used to resolve ambiguities in the echo
arrival angle determinations, and to provide a measure of the error in arrival
angle determination; again, the groundwave provides the reference phase at each
of these antennae.

This direction finding system is often referred to as an "all sky" system,
since it employs dipoles mounted $\lambda/4$ above ground, and "sees" echoes from a
"slab" of the meteor region centered on the transmit/receiver system, and
approximately 200 km in diameter.

MULTI-STATION SYSTEMS

The multi-station technique developed at Jodrell Bank (GILL and DAVIES,
1956; DAVIES, 1957) for the measurement of the orbits of individual meteors has
seen application in the determination of wind shears on individual trails, and
the statistics of average wind shear have been interpreted in terms of
atmospheric turbulence (e.g., ROPER, 1966).

Since radio reflection from meteor trails is specular, spacing receiving
sites allows simultaneous observation of different parts of individual meteor
trails. The original Adelaide system is described in WEISS and ELFORD (1963),
and an updated system detailed by ROPER (1972) is sketched in Figure 2.

The geometry is such that maximum reflection point separation is half the
maximum separation between receiving sites. Results from the Adelaide system of
Figure 2 have shown that very few trails are more than 10 km long. Reflection
point separations of less than one half a Fresnel zone length will be highly
correlated - thus shears can be measured over trail lengths from some 0.5 to 10
km. Station separations of approximately 7 km represent the optimum for orbit
and wind shear determinations.

ECHO ACCEPTANCE CRITERIA

In addition to the "delay" before echo sampling begins, certain acceptance
criteria need to be applied to each echo. Those applied to the Atlanta system
are here used for illustration, but are appropriate for any system.

1. The doppler must be coherent for at least 1 1/2 cycles.
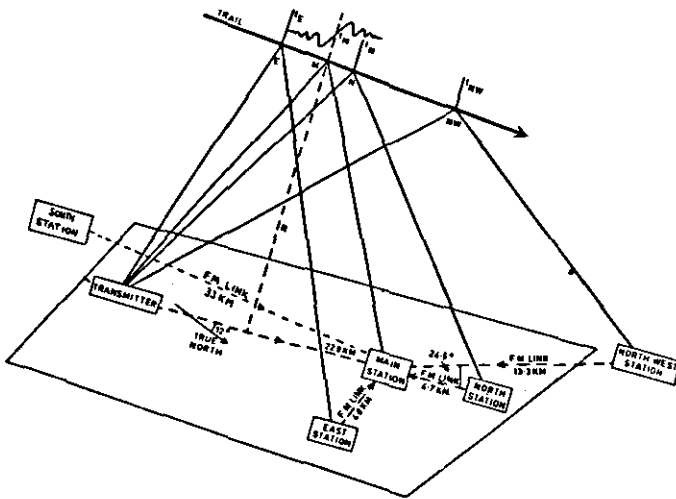2. The error in echo arrival angle (measured for each echo < 1°.

Figure 2. Measurement layout.

3. Range should be consistent (within instrumental accuracy).

FURTHER ACCEPTANCE CRITERIA

Because of the diurnal variation in echo rate, it may actually be necessary to discard a portion of the valid echoes measured in high rate hours in order not to bias wind determinations toward these high rate periods. Similarly great care needs to be exercised in analyzing data in which significant gaps occur, although there are interpolation schemes available which can be applied to minimize the abiasing due to data gaps. A recent paper by CRARY and FORBES (1983) considers in detail both the requirements of measurement accuracy and frequency of observational data required to satisfactorily delineate tidal winds.

While any reasonably well designed radar will introduce short term instrumental errors that are significantly less than the "error" due to the presence of random wind components in any mean wind or tidal observation, the importance of system calibration cannot be overemphasized; such calibration (provided by a phase coherent low level source in the neighborhood of the antenna array) can readily be performed automatically.

TYPICAL RADAR INSTALLATIONS

Reference has already been made to the Atlanta, Adelaide, and CNET radars, which are fundamentally CW systems. This section will therefore concentrate on pulsed radars.

(a) The Stanford System

This system has been described at length by NOWAK et al. (1970). It is a long pulse (280 μs) radar, with 28 bit phase encoding of the pulse producing ±2 km range resolution. Echo height was determined by using an amplitude comparison of signals received by two antennae, one above the other. Azimuth was assumed to be the pointing azimuth of each yagi array.

An updated version of this system, using interferometry to deduce both elevation angle and azimuth (see RUDMAN et al., 1970) continues to be operated by the Electrical Engineering Department of the University of New Hampshire at Durham.

Figure 3 is a block diagram of the latest version of this system, currently being used by the Radio Atmospheric Science Center of the University of Kyoto, Japan (TSUDA, 1982).

(b) The Soviet VETA Radar

This radar (Figure 4) was developed at the Kharkov Institute of Radio Electronics (KASCHEYEV and NECHITAILENKO, 1974). Unfortunately, in common with the majority of Russian meteor wind radars, height is not measured.

The Hydrometeorological Service of the U.S.S.R. operates the only national observational network in the world, with stations from Kuhlungsborn in the west to Kharbarovsk in the east, from Heiss Island in the north, to Volgograd in the south. The Soviets have made significant contributions to our understanding of mesopause level dynamics; their contribution would be greatly enhanced, however, by the use of height finding systems at each of their stations.

(c) The Proposed Georgia Tech Radio Meteor Observatory

I am including the proposed Georgia Tech Radio Meteor Observatory (a three station system) only as an illustration of a system which can continuously measure the variation with height of mesopause level winds; provide estimates of the rate of dissipation of turbulent energy from the shears measured on individual trails; and also measure the orbits of individual meteors.
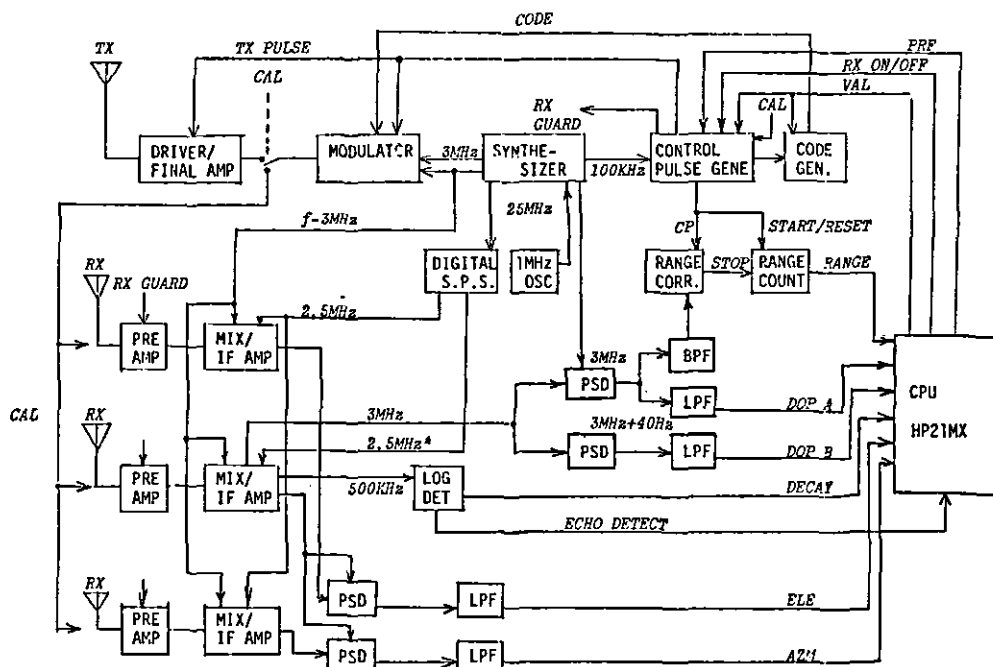

Figure 3. Kyoto University meteor wind radar block diagram.
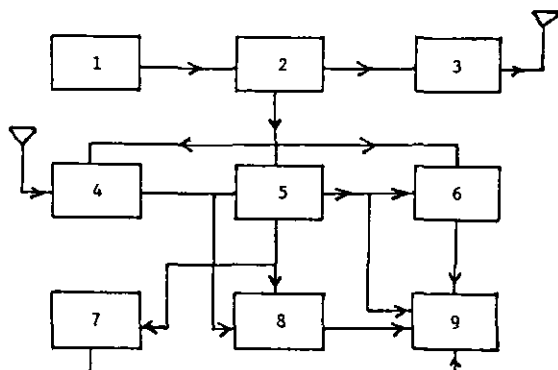
130

RUSSIAN VETA RADAR

Figure 4. 1 - electronic clock; 2 - synchrogenerator;
3 - pulse coherent transmitter; 4 - reciever;
5 - preliminary processing device; 6 - digital
range meter (R-meter); 7 - amplitude-code converter;
8 - HF digital pulse phasemeter; 9 - on-line computer
or (and) input device of a communication link.

At the main station (Figure 5), there is a pulsed radar feeding a dipole antenna, with 5 receivers fed from the direction finding antenna array of Figure 1. The system is computer controlled using an already available minicomputer. A modern system being built could profitably use microprocessor control. With this type of control, system operation parameters, such as prf, pulse width, peak power are software controlled. For example, in the "patrol" mode, 10 μsec pulses at an unambiguous prf are transmitted, and received on wideband (100 kHz) receivers; once a meteor echo is recognized, the system changes to "record" mode, with a longer pulse (266 μsec) and a higher prf, and the receivers are switched to a narrow passband, with coherent offset frequency injected into each receiver for doppler and arrival angle determination.

Each outstation (Figure 6) continuously feeds data to the main station via a 1200 baud two-way UHF link. Thus the presence of an echo at any of the three stations can initiate recording. All recording is done on magnetic tape at the main station. All echoes producing sufficient data at the main station will be used for wind determination; those echoes producing adequate data from one or both outstations and the main station will contribute to the atmospheric turbulence statistics; and those echoes producing adequate data from all 3 stations will be used in the orbit survey.

The computer will perform online data processing. The major portion of the data reduction and analysis will be performed offline, using as input the magnetic tape output of the system.

DATA REDUCTION AND ANALYSIS

For determination of the height/time wind variations one needs to measure

1. Line of sight drift velocity (both magnitude and direction).
2. Echo height.
3. Time of occurence of the echo (year, day of year, hour, minute, second).

Several methods of wind analysis assume the vertical wind to be zero, and simply

PROPOSED GEORGIA TECH RADIO METEOR SYSTEM

Figure 5. Main station transmitting and receiving system.

project the line of sight component onto the horizontal plane. This approxi-
mation is not serious, but is not made in at least two analysis methods, those
of GROVES (1959) and SALBY (1978). Groves' analysis has been used by several
groups; it produces mean vertical winds which are of the order of meters per
second, much larger than expected. However, it does appear that the _sign_ of
these vertical winds is correct. This is an area to which MST VHF radars can
make a considerable contribution, since they are able to measure vertical winds
directly.

Planetary waves and tides can be studied by even low-powered systems.
Since the setup time for tides is of the order of days, the average tidal wind
components for 5 to 10 day intervals are best compared to theoretical models.

COOPERATIVE OBSERVATIONS

For the decades of the 50's and 60's, so much was being learned from
observations of mesopause level dynamics that individuals made highly
significant contributions with very little cooperation between stations.

However, at the International Symposium on Waves in the Upper Atmosphere
held in Toronto in 1970, Colin Hines proposed that in order to further our
understanding of motions at these altitudes, a cooperative approach was
necessary. This suggestion was taken up by the International Association of
Geomagnetism and Aeronomy Working Group on Meteors, resulting in GRMWSP, the
Global Radio Meteor Wind Studies Project. A subsequent proposal from the
Incoherent Scatter Group of the International Union of Radio Science resulted in
CTOP, the Cooperative Tidal Observations Program, which used incoherent scatter

132

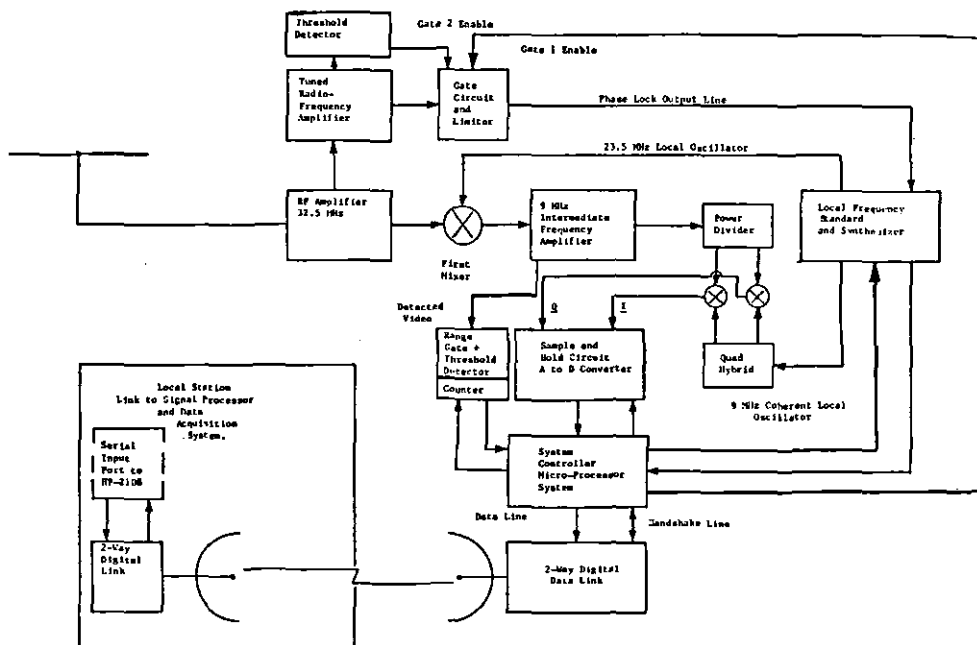PROPOSED GEORGIA TECH RADIO METEOR SYSTEM

Figure 6. Outstation receiver and data handling system (2 of).

tidal data to extend the height range of observations. This cooperation resulted in several papers being published in the August, 1978 issue of the Journal of Atmospheric and Terrestrial Physics.

More recently (1982), the Middle Atmosphere Program project GLOBMET (Global Meteor Observations System – for details see MAP Handbook No. 7, page 20) has further formalized these cooperative efforts.

In concluding, I should add that meteor wind radars are just one of the tools available for the determination of mesopause level dynamics. A truly global picture of this region of the atmosphere will only be forthcoming if all techniques able to provide data on this region – e.g., partial reflections, incoherent scatter radar, MST radars, as well as satellites and rockets are used.

ACKNOWLEDGEMENTS

REFERENCES

Crary, D. J. and J. M. Forbes (1983), On the extraction of tidal information from measurements covering a fraction of a day, Geophys. Res. Lett., 10, 580-582.

Davies, J. G. (1957), Radio observations of meteors, in Advances in Electronics and Electron Physics, 9, 95-128, Academic Press, Inc., New York.

Dolas, P. M. and R. G. Roper (1982), Prevailing wind in the meteor zone over Atlanta and its association with midwinter stratospheric warming, J. Atmos. Sci., 38, 182-188.

Gill, J. C. and J. G. Davies (1956), A radio echo method of meteor orbit determination, Mon. Not. Roy. Astron. Soc., 116, 105-113.

Glanz, F. H. and R. J. Rudman (1971), An azimuth measuring system for meteor trails radar, Trans. IEEE on Geosci. Electronics, 9, 1, 56-62.

Groves, G. V. (1959), A theory for determining upper atmosphere winds from radio observations on meteor trails, J. Atmos. Terr. Phys., 16, 344-356.

Kascheyev, B. L. and V. A. Nechitailenko (1974), The meteor wind radar I. Problems and principles of design, Radiotechnika, N 31, Kharkov (in Russian).

Nowak, R., E. M. North and M. S. Frankel (1970), The Stanford meteor-trails radar Mark II, Tech. Rep. AFCRL-70-0365, Air Force Cambridge Research Labs.

Roper, R. G. (1966), Atmospheric turbulence in the meteor region, J. Geophys. Res., 71, 5785-5792.

Roper, R. G. (1972), Radio Meteor Winds in the Southern Hemisphere, Chapter 8 of Thermospheric Circulation, ed. Willis L. Webb, MIT Press.

Roper, R. G. (1975), The Georgia Tech radio meteor wind facility, Final Technical Report, National Science Foundation Grant No. GA26626, Georgia Institute of Technology.

Robertson, D. S., D. J. Lilly and W. G. Elford (1953), Measurements of winds in the upper atmosphere by means of drifting meteor trails, I., J. Atmos. Terr. Phys., 4, 255.

Rudman, R. J., F. H. Glanz, D. D. Frost and R. R. Clark (1970), An interfero-meter system for the measurement of radar echoes from meteor trails, Tech. Rep. AFCRL-70-5088, Air Force Cambridge Research Labs.

Salby, M. L. (1978), Radio meteor wind analysis using finite element approxima-tion, J. Atmos. Terr. Phys., 40, 636-656.

Tsuda, J. (1982), Kyoto meteor radar and its application to observations of atmospheric tides, Ph.D. Thesis, Radio Atmospheric Science Center, Kyoto University, Japan.

Weiss, A. A. and W. G. Elford (1963), An equipment for the combined geophysical and astronomical measurements of meteors, Proc. I.R.E. Australia, 24, 197-293.

FURTHER BACKGROUND

Barnes, A. A. and J. J. Pazniokas (1968), Proceedings of the Workshop on Methods of Obtaining Winds and Densities from Radar Meteor Trail Returns, Special Report No. 75, AFCRL-68-0228, Air Force Cambridge Research Laboratories.

Journal of Atmospheric and Terrestrial Physics (1955), (Special Supplement) 2.

McKinley, D. W. R. (1961), Meteor Science and Engineering, McGraw Hill Book Company, Inc., New York.

Webb, W. L. (1972), Thermospheric Circulation, MIT Press, Cambridge, Massachusetts.

APPENDIX I.   CNET System Bibliography.

(1)   Belbeoch G., Project d'etude CNET-DRME, 1274/RSR/AS (1970).

(2)   Glass M., Revah I., Spizzichino A., Radars meteoriques mobiles pour l'etude des vents dans la haute atmosphere (Note tech. 1969) EST/RSR/4.

(3)   Revah I., Etude des vents de petite echelle observes au moyen des trainees meteoriques (These de Docteur es Sciences, 1968) GBI/NT/77.

134

(4) Spizzichino A., Etude des interactions entre les differentes composantes du vent de la haute atmosphere (These de Docteur es Sciences, 1969) Annexes 22 er 23.

(5) Spizzichino A., Delcourt J., Geraud A. et Revah I., A new type of continuous wave radar for the observation of meteor trails, Proc. IEEE 53 1084-1086 (1965).

(6) Bannerot G., Etude d'un radar meteorique a frequences alternees (memoire de Diplome d'Etudes Superieures, 1971), Laboratoire de Physique de l'Ecole Normale Superieure (Paris).

(7) Belbeoch G., Depouillement automatique d'echoes meteorique (Note tech. a paraitre, 1972) CNET/RST.

(8) Deberne, D., Glass, M., Boulett B., et Belbeoch G., Systeme d'Exploitation par calculateur des radars meteoriques du CNET (Note techn. 1972) GRI/NTP/97.

11.   INCOHERENT SCATTER RADAR STUDIES OF THE MESOSPHERE

J. D. Mathews

Department of Electrical Engineering and Applied Physics
Case Institute of Technology
Case Western Reserve University
Cleveland, OH 44106

ABSTRACT

All current incoherent backscatter radars can make mesospheric or D region measurements under at least some conditions. The conditions under which measurements are possible are derived and shown for each radar. Conditions examined include overall system sensitivity, electron and ion spectral line widths and power distributions, and time/height averaging effects. The radar equation is derived and calibration of the radar system is discussed. Total power only and combined total power and ion line spectral measurements are described in terms of ease of use, applicability to various radars, and parameters of aeronomic interest derivable from the measurements.

INTRODUCTION

The use of UHF incoherent scatter radars (ISR) for study of the mesosphere has reached sufficient maturity that information regarding current measurement capabilities as well as the potential for future enhanced and/or multitechnique measurements should be made available to the community of middle atmosphere researchers. This paper, which is tutorial in nature, is an attempt to provide this information.

Use of the ISR for middle atmosphere studies has evolved in three major stages which are related to enhancements in technology. The first stage involved total power or electron concentration (N_e) measurements using pulsed techniques yielding height resolutions of 3-6 km and a minimum observable N_e of about 600-1000 el/cc. Results were reported by LALONDE (1966, 1967) and THOME and WAGNER (1971) using Arecibo. ARMISTEAD et al. (1972) give Millstone *Hill results while TAYLOR (1975) describes solar flare effects on the 80-140 km* ionosphere at Malvern (UK). Taylor also compares the Malvern results with Arecibo flare effect observations.

The second stage of development was characterized by introduction of Barker code techniques (IOANNIDIS and FARLEY, 1972; GRAY and FARLEY, 1973; RIHACZEK, 1969) which permitted height resolutions sufficient to fully resolve the D region ledge while maintaining high per pulse energy. IOANNIDIS and FARLEY (1974) give Arecibo N_e results with 0.9 km height resolution and also derive the corresponding effective recombination coefficients. Further second stage Arecibo results are given by TROST (1979) who also gives a thorough analysis of errors. The minimum detectable N_e remained essentially the same as in the first stage.

Both the first and second stages of development were characterized by lack of computer power sufficient to fully analyze all data. This gave rise to use of various signal sampling and processing techniques which were somewhat less than optimum in terms of resultant signal-to-noise ratio but allowed for more efficient utilization of available computer resources (see FARLEY, 1969). SECHRIST (1974) compares various techniques for D region N_e measurements and summarizes the incoherent scatter measurements reported up to then.

The third and current stage of development involves both power (N_e) and spectral measurements and has been made possible by significant advances in computing power (often in the form of array processors), and by sophisticated feeds and low-noise receiver "front-ends". MATHEWS (1976) reports Arecibo 80-100 km wind measurements found using pulse-to-pulse correlation methods. HARPER (1978) reports full D-region "ion line" spectral measurements. FUKUYAMA (1981) reports time and height variations of ion line spectral width and power. Combined power and spectral width have been used to determine, subject to scattering theory, various atmospheric/ionospheric parameters including winds, temperature, pressure, ion (negative and positive) concentration, and perhaps ion mean masses, (TEPLEY and MATHEWS, 1978; GANGULY et al., 1979; MATHEWS et al., 1981; TEPLEY et al., 1981). MATHEWS et al. (1982) reports a minimum detectable N_e of 100 el/cc with a total systematic plus five standard deviation statistical error range of 50-180 el/cc for the lower D region at Arecibo.

In this paper we examine, in some detail, the methods, results, and possible extensions of the third and current stage of development in ISR measurements of the mesosphere. We begin in the next section by examining the form of the radar equation appropriate to ISR work. The section entitled "the power profile" concerns total power or N_e measurements alone. In the section "combined total power and ion line power spectral measurements" we discuss the consequences of combined total power and power spectral measurements. The last section concerns recent results, conclusions, and possibilities for future extended and/or multitechnique measurements.

THE RADAR EQUATION

In all ISR experiments we are ultimately interested in relating received power to transmitter power, removing height or range variations not associated with the scattering medium, and finding the scattering cross-section (total or as a function of frequency) per unit volume. The required relationship is known as the radar equation and since this equation is vital to data reduction, a brief development of it follows.

The far-field transmitter electric field incident on the electrons in the illuminated volume accelerates each electron which then reradiates or scatters proportionally to the electric field magnitude (BEKEFI, 1966; BERNE and PECORA, 1976) and subject to the "random" forces imposed on it by the surrounding medium. The power (per unit solid angle) scattered by each electron is then proportional to the incident energy flux (Poynting vector magnitude) times the electron scattering cross-section (classical electron radius (r_e) squared; see REITZ and MILFORD, 1967, section 19-3). However if a plane EM wave illuminates a uniform electron gas the net (macroscopic) scattered power is zero. This result easily follows from the geometry of Figure 1 if we divide the uniform scattering medium into thin sheets oriented perpendicular to $\underline{k} = \underline{k}_i - \underline{k}_s$. All electrons in each sheet radiate in phase along $\underline{k}$ and for each sheet we choose a second sheet located distance $d = \pi/|\underline{k}_i - \underline{k}_s| (= \lambda/4 \sin (\theta/2)$; see Figure 1) away that contributes radiation which would exactly cancel that of the first sheet.

For non-zero net scattered power to occur the medium must exhibit concentration fluctuations. In particular maximum scattering or minimum cancellation occurs from fluctuations characterized by wavenumber $k = 2\pi/2d (=|\underline{k}_i - \underline{k}_s|)$. This is the classical Bragg scattering condition which may also be derived from quantum electrodynamics by invoking momentum and energy conservation of a photo (FEYNMAN et al., 1965) interacting with or creating a disturbance (fluctuation) in the illuminated medium. Then momentum and energy conservation yield, respectively,
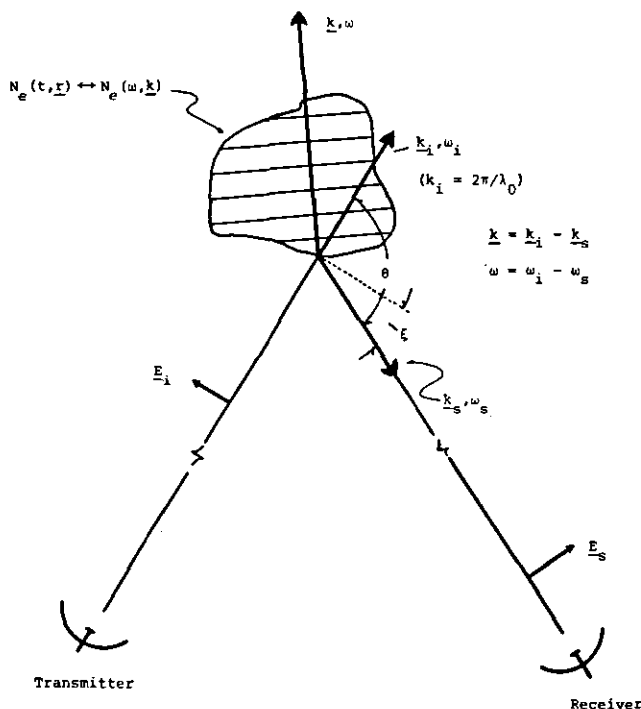
Figure 1. Scattering geometry with k vectors to scale for the angles involved. $E_{i,s}$ are the incident and scattered electric field intensities respectively.

$$h\underline{k} = h\underline{k}_i - h\underline{k}_s$$

$$h\omega = h\omega_i - h\omega_s \tag{1}$$

where h is Planck's constant and $h\underline{k}$ and $h\omega$ characterize the momentum and energy of the disturbance.

The net amount of scattered power is then proportional to the power spectrum of electron concentration fluctuations evaluated at $\underline{k},\omega$. This result can be expressed in terms of σ, a differential scattering cross-section per unit volume per Hertz, where

$$\sigma(\omega_i + \omega) = r_e^2 \sin^2 \xi \ <|N_e(k,\omega)|^2> \ / \ \Delta V \Delta T \tag{2}$$

with ξ indicated in Figure 1, $N_e(\underline{k},\omega)$ is the double Fourier transform of $N_e(\underline{r},t)$ evaluated on volume ΔV and over the time interval ΔT. The brackets indicate ensemble averaging. It can be demonstrated (KITTEL, 1958; chapter 25, example 25.4) that $<|N_e(\underline{k},\omega)|^2>$ is proportional to $N_{eo}\Delta V \Delta T$ where N_{eo} is the steady state electron concentration. The detailed calculations of $<|N_e(\underline{k},\omega)|^2>$ are usually based on application of the generalized Nyquist Theorem (BEKEFI, 1966) to fluctuations described by the combined Maxwell equations and the plasma state equations. Results are given by DAUGHERTY and FARLEY (1960), TANENBAUM (1968), EVANS (1969), and many others. These results may be parameterized by rewriting (2) as

138

$$\sigma(\omega_i + \omega) = N_{e0} \; 4\pi \; r_e^2 \; \sin^2 \xi \; \sigma_n(\omega) \tag{3}$$

where σ_n is the normalized scattering coefficient per electron $(0 \leq \sigma_n \leq 1)$ and the 4π "converts" σ to an isotropic scattering cross-section per unit volume.

We are now able to derive the backscatter $(\xi = 90°)$ radar equation based simply on standard antenna terminology. Let $G(\theta,\phi)$ be the antenna gain over a lossless isotropic radiator (KRAUS, 1950; section 2.15) where θ is measured from the beam central axis and ϕ is the corresponding azimuthal angle. The energy flux at range r (in the far-field) is just

$$F_i(\theta,\phi) = \frac{P_T \; L \; G(\theta,\phi)}{4\pi r^2} \tag{4}$$

where P_T is the transmitter power and L (≤ 1) is the coefficient of transmitting system efficiency. The scattered energy flux (per unit scattering volume per Hertz) at the receiver is just

$$F_s(\theta,\phi) = \frac{F_i(\theta,\phi) \; \sigma(r,\omega)}{4\pi r^2} \tag{5}$$

and the effective aperture (or collecting area; this is a measure of the capability of an antenna system to convert incident energy flux to power available to the receiver) of the antenna is $A_e(\theta,\phi) = \lambda^2 G(\theta,\phi)/4\pi$ (KRAUS, 1950; STUTZMAN and THIELE, 1981, section 1.10). Then received power (per unit scattering volume) is just

$$P_s{}'(r,\theta,\phi;\omega) = \frac{P_T L \lambda^2 G^2(\theta,\phi) \sigma(r,\omega)}{4\pi \cdot (4\pi r^2)^2} \tag{6}$$

where we see immediately that for σ a fixed scattering cross-section (say, an aircraft) the received power is proportional to r^{-4}. However, we find the total received power per unit frequency by integrating over the scattering volume. This yields

$$P_s(r;\omega) = \frac{P_T L \lambda_0^2 \sigma(r,\omega) c\tau}{128\pi^3 r^2} \int_0^{2\pi} \int_0^\pi G^2(\theta,\phi) \; \sin\theta d\theta d\phi$$

where we have assumed the differential volume to be $0.5 \; c\tau \; r^2 \sin\theta d\theta d\phi$ with τ the pulse length and c the speed of light and that $\sigma(r,\omega)$ is essentially constant over the contributing volume. Finally, for azimuthal beam symmetry and using (3) (with N_e replacing N_{e0}), we find that

$$P_s(r;\omega) = \frac{P_T L N_e(r) \sigma_n(r;\omega) c\tau \lambda^2 r_e^2}{16\pi \; r^2} \int_0^\pi G^2(\theta) \; \sin\theta d\theta \tag{7}$$

This is the usual form of the backscatter radar equation. It is possible by use of radio astronomy based pattern measurements to directly determine the value of the integral (known as the backscatter integral I_{bs}) in (7) (BREAKALL and MATHEWS, 1982). Alternatively, plasma line and/or ionosonde measurements may be used to indirectly determine I_{bs}. These techniques are discussed in MATHEWS et al. (1982).

We can rewrite (7) in a form that best allows comparison of various mono-static ISR's. To do this let

$$G(\theta) = G_0 P_n(\theta) \tag{8}$$

where P_n is the normalized pattern and G_0 is peak gain. Then the back-scatter integral I_{bs} becomes

$$I_{bs} = 2G_0 \epsilon_r k_{bs} \tag{9}$$

where

$$k_{bs} = \frac{\int_0^\pi P_n^2(\theta) \sin \theta d\theta}{\int_0^\pi P_n(\theta) \sin \theta d\theta} \tag{10}$$

but the denominator of (10) is just $2\epsilon_r/G_0$ (STUTZMANN and THIELE, 1981) where directivity and gain are related by $G_0 = \epsilon_r D$ with $\epsilon_r (\leq 1)$ the ohmic efficiency coefficient. Using (9) the radar equation becomes

$$P(r;\omega) = \frac{K_{sys} P_T N_e(r) \sigma_n(r;\omega)}{r^2} \tag{11}$$

$$K_{sys} = \frac{c\tau\epsilon_r \lambda_0^2 G_0 L k_{bs} r_e^2}{8\pi}$$

The quanity k_{bs} is of order 0.4 with $k_{bs} = 0.41 \pm 0.01$ for the Arecibo 430 MHz system (BREAKALL and MATHEWS, 1982) and $k_{bs} = 0.38$ for a 10 dB taper parabolic dish (EVANS, 1969).

THE POWER PROFILE

The so-called total power or power profile measurements still form the principal use of the ISR and much information concerning the D region can be obtained from these measurements alone (MATHEWS et al., 1982). For example, the total power information can be converted to $N_e(z,\chi)$ (where z is height and χ is solar zenith angle) and may then be used to study the photochemistry of the region. Clearly however the quality of $N_e(z,\chi)$ depends on how well the signal power profile has been extracted from the signal-plus-noise power profile and on how well the conversion from signal power to N_e has been accomplished. We will now consider these two levels of processing.

Figure 2 shows what might be a typical "averaged" power profile determined using short or coded pulses plus calibration sequence. The base level is just the averaged noise power ("available" in the matched load resistance)

$$P_n = k_b T_{sys} B \tag{12}$$

where T_s is the system temperature, k_b is Boltzmann's constant and B is the system noise equivalent bandwidth where

$$B = \int_0^\infty S_n(f) \, df \quad . \tag{13}$$

$S_n(f)$ is the normalized system (filter) power transfer function. The calibration source temperature is assumed to be a "standard" and thus forms the basis for relating the relative power profile to the radar equation and thus to
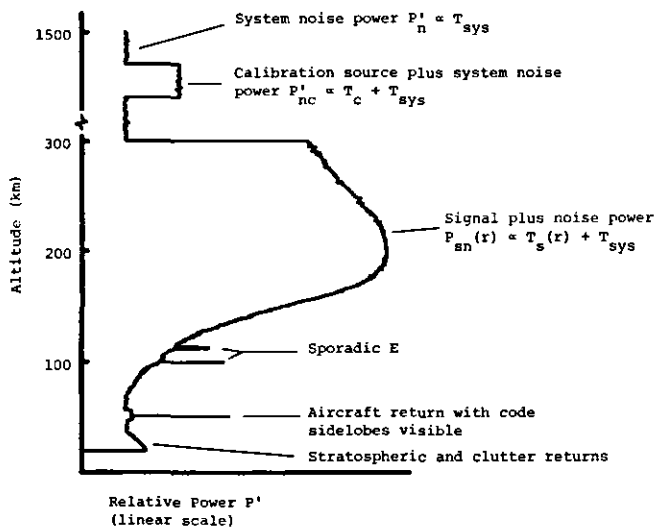
Figure 2. Pictorial representation of a
power profile produced by averaging over
say 2000 radar pulses. Features shown
include the calibration sequence plus
ionosphere, interference, and clutter
returns.

absolute scatter power. Specifically it is clear from Figure 2 that the signal
temperature is

$$T_s(r) = \frac{P'_{sn}(r) - P'_n}{P'_{nc} - P'_n} \; T_c \tag{14}$$

where P'_s is relative power with subscript s referring to signal-plus-noise
(sn), noise only (n), or noise plus calibration source noise (nc). The signal
temperature, $T_s(r)$, is related to signal power at the antenna load and thus to
the radar equation (11) by

$$P_s(r) = k_b T_s(r) B \tag{15}$$

assuming matching between antenna and load.

As indicated earlier the P'_s quantities in (14) are averaged over many
separate radar pulse returns. This averaging is necessary since the signal,
noise, and calibration source noise voltages measured across the antenna load
are Gaussian random processes (assuming no interference). Thus single voltage
samples (in phase and in quadrature) at a single height gate have little meaning
until they are squared, added, and averaged over many radar pulses (individual
experiments) forming one of the quantities P'_s seen in Figure 2 or equation
(14). If we assume a pulse repetition rate of 100 sec^{-1} then in 1/2 hours,
assuming no time "overhead", we can average together $N = 1.8 \times 10^5$ power
samples at each height gate. The resultant signal error level (assuming
stationarity) is

$$\frac{\Delta P_s}{P_s} = \frac{1 + P_n/P_s}{\sqrt{N}} \tag{16}$$

where ΔP_s is one standard deviation and we have assumed sampling in phase and in quadrature as well as "ideal" (full multiples) detection (FARLEY, 1969). Since in Figure 2 noise and noise plus calibration noise regions extend unchanged over many height gates we also assume use of height averaging over, say, 200 height gates, to significantly increase the number of samples used to estimate P_n' and P_{nc}' in (14). We are thus justified in excluding errors in estimating P_n' or P_{nc}' from our estimate of error in P_s given by (16). These errors would be of an order of $\sqrt{N_g}$ smaller than ΔP_s where N_g is the number of "noise height gates" averaged.

Using the radar equation, as given in (11) and equation (12), along with various system parameters it is possible to estimate the single pulse signal-to-noise ratio (SNR) of various upper VHF and UHF backscatter radars. This is done in Table 1 using typical parameter values (e.g., $k_{bs} = 0.4$; $\sigma_n = 1/2$) and for 600 m height resolution which we assume to be achieved using a 13 baud Barker code/decoder and thus 52 μsec long pulses and a 250 kHz receiver bandwidth. The resultant single pulse SNR's are small but averaging over 1.8×10^5 samples in time and 5 height gates yields from (16) the statistical uncertainties given in Table 2. The Table 1 results are however misleading for the upper UHF radars because significant scattered power lies outside the assumed 250 kHz bandwidth for these radars. This problem will be addressed after discussing interference.

TABLE 1.   COMPARISON OF UHF BACKSCATTER RADARS

Site	λ (m)	P_T(MW)	G_0 (dbi)	T_{sys} (°K)	SNR for 1 pulse[a,b]	
					$N_e = 10^4$/cc; z=95 km	$N_e = 10^3$/cc; z=80 km
Arecibo (430 MHz)	0.698	2.5	61.1	120	3.00	.42
EISCAT (224 MHz)	1.338	5.0	43.4	160	.28	.04
EISCAT (934 MHz)	0.321	2.0	48.1	120	.03	.003
Millstone (440 MHz)	0.682	2.0	46.5 (150 ft)	120	.07	.01
Millstone (440 MHz)	0.682	2.0	49.8 (220 ft)	120	.15	.02
Sondrestrom (1290 MHz)	0.232	5.0	49.6	110	.05	.007

Parameters Used:

$k_{bs} = 0.4$, $\tau = 52$ μsec, $L = 0.75$, $\epsilon_r = 0.97$, $B = 250$ kHz, $\sigma_n = 1/2$

Caution:

a) Assumes sampling every four μsec yielding 600 m range resolution if 13 baud Barker code can be successfully decoded.

b) Use of $\sigma_n = 1/2$ assumes that all electron line power is within B bandwidth. This is not necessarily true; see Table 4.

TABLE 2. TYPICAL STATISTICAL UNCERTAINTIES[1]

SNR (single pulse)	$\Delta P/P^2$	$\Delta P/P^{2,3}$
1	.0047	.0021
.1	.026	.012
.01	.24	.11

[1] Assuming noise power is much more accurately estimated than signal power (see text).

[2] Assumes 1.8×10^5 samples. i.e., a pulse rate of 100 sec^{-1} for 1/2 hour.

[3] Assumes averaging over 5 height gates with SNR constant.

As alluded to previously interference effects can seriously affect estimates of P_s in the small SNR regime characteristic of D region measurements and should therefore be eliminated where possible. While the best method of treating interference is prevention several post occurrence treatment methods are possible. One method involves retaining separate power profiles averaged over, say, 20 seconds (or 2000 pulses) and examining each intermediate profile for interference effects from, say, aircraft or satellite returns. This examination can be either semi-manual or automatic but should be based on running statistical predictions of what power fluctuations are expected. An "alarm level" of five standard deviations is probably sufficient. Other interference discriminators can include tests for long correlation time returns in the manner of ARMISTEAD et al. (1972) and IOANNIDIS and FARLEY (1974). It is also possible to form spectral estimates over narrow unaliased bandwidths using FFT methods and then subtracting power in narrow "lines" from the total power. The FFT method is basically that used to estimate the "ion line" spectrum and is discussed in the next section. The use of a similar method to account for unwanted signal components is discussed by SATO and WOODMAN (1982).

All needed signal averaging in both height and time should be done to the power profiles. This assures correct data treatment in a statistical sense. Once the averaged power profile is determined the radar equation can be used to find N_e. In doing this conversion we must check that filter bandwidths are wide enough to contain not only the ion line but also the total of the electron line if there is significant electron line power. This is important in the lower D region where the Debye length can be sufficiently large that a significant fraction of the scattered power is in the electron line. However, MATHEWS (1978) and MATHEWS et al. (1982) note that the collision dominated electron line is considerably narrower than that expected from electron mean thermal speed considerations alone.

To quantify the comments in the previous paragraph we note (FARLEY, 1966) that (for $1 \leq T_e/T_i \leq 3$ and no negative ions) the total normalized scattering coefficient per electron (see equation (3)) is

$$\sigma_n = \frac{\alpha_e^2}{1 + \alpha_e^2} + \frac{1}{(1 + \alpha_e^2)(1 + T_e/T_i + \alpha_e^2)} \tag{17}$$

where $\alpha_e = 4\pi\lambda_D/\lambda_0$ and λ_D [$\approx 69 (T_e/N_e)^{1/2}$ m] is the Debye length with λ_0 the radar wavelength. In (17) the first and second terms on the right hand side indicate the electron and ion line contributions, respectively, to the total scattered power. Thus, for example, when $\alpha_e^2 \gg 1$ ($\lambda_D \gg \lambda_0$) $\sigma_n = 1$ which is contributed solely from the electron line. While for $\alpha_e^2 \ll 1$ ($\lambda_D \ll \lambda_0$)

$$\sigma_n (\lambda_D \ll \lambda_0) = \frac{1}{1 + T_e/T_i} \tag{18}$$

which is contributed from the ion line. Note that when the ratio of electron and ion temperatures (T_e/T_i) is 1 then $\sigma_n = 1/2$.

The distribution of ion line and electron line total power levels is indicated for the various radars of Table 1 in Table 3 for $N_e = 10^2$, 10^3, 10^4 el/cc. From Table 3 we note that at 10^3 el/cc the two upper UHF radars have considerable electron line power and thus that wide filters must be used to encompass the electron component of the spectrum. Table 4 gives electron line thermal widths (W_{te}) and approximate collision dominated widths (W_{ce}) for ψ_e (the normalized electron collision frequency, DOUGHERTY and FARLEY, 1963; MATHES, 1978) greater than one. We see from Tables 3 and 4 that the Sondrestrome and EISCAT (934 MHz) radars probably cannot measure N_e below $\sim 10^4$ el/cc because significant power lies outside any practical filter band-widths. Any increase in filter bandwidths would allow a commensurate increase in noise power; however some small compromise might be possible.

We also note that significant electron line power implies that any pulse coding scheme used to achieve high height resolution is in fact not working. For the 52 μsec, 13 baud Barker code assumed for Table 1 calculations the height resolution would become 7.8 km not 0.6 km. The criterion for successful applicaton of the code is that the medium does not decorrelate in less than the 52 μsec code length. This corresponds to a maximum signal bandwidth of about 20 KHz even though the full 250 KHz is required to actually get independent measurements every 600 m.

Given that essentially all scattered power is represented in the signal power profile $N_e(z)$ may be found from (11), the radar equation, where K_{sys} has been determined from, say, plasma line or ionosonde measurements (MATHEWS et

TABLE 3. RATIOS R_{ei} OF ELECTRON LINE TO ION LINE TOTAL POWER FOR VARIOUS RADARS

Site	$R_{ei} = \alpha_e^2 \cdot (2 + \alpha_e^2)$ and $T_e = T_i = 200°K$		
	$N_e = 10^4$ el/cc $\lambda_D = 1$ cm	$N_e = 10^3$ el/cc $\lambda_D \simeq 3.1$ cm	$N_e = 10^2$ el/cc $\lambda_D \simeq 9.8$ cm
Arecibo	.06	0.72	15.8
EISCAT (224 MHz)	.02	0.18	2.4
EISCAT (934 MHz)	.31	5.1	242.
Millstone	.07	0.74	16.6
Sondrestrom	.36	6.1	312.

TABLE 4. ELECTRON LINE WIDTHS FOR THE VARIOUS RADARS

Site	Electron Line Thermal Width[1] W_{te}(kHz)	ψ_e (80km)[2,3]	Electron Line[5] Collisional Width at 80 km W_{ce}(kHz)	ψ_e (70 km)[2,4]	W_{ce}(kHz)[5] at 70 km
EISCAT (224 MHz)	265	1.8	150	11	24
Arecibo (430 MHz)	510	.95	510	5.8	88
Millstone (440 MHz)	520	.90	520	5.4	96
EISCAT (934 MHz)	1100	.43	1100	2.6	425
Sondrestrom (1290 MHz)	1530	.03	1530	0.2	1530

[1] $W_{te} \simeq 355/\lambda_0$ (m) kHz with $T_e = 260°K$

[2] $\psi_e = \nu_{en}/\sqrt{2}\,k_s U_e$; $k_s = 4\pi/\lambda_0$, λ_0 = radar wavelength, $U_e = 6.28 \times 10^4$ m/sec = electron thermal speed

[3] $\psi_e = 1.35\,\lambda_0$ for $\nu_{en} = 1.5 \times 10^6$ sec^{-1}

[4] $\psi_e \simeq 9\,\lambda_0$ for $\nu_{en} = 1.0 \times 10^7$ sec^{-1}

[5] $W_{ce} \simeq W_{te}/\psi_e$ for $\psi_e > 1$ (MATHEWS and TANENBAUM, 1981)

al., 1982; BREAKALL, 1983). When inverting (11) both N_e and σ_n are unknown but since σ_n is a function of N_e (and T_e; see equation (17)) an iterative procedure is sufficient to find mutually consistent N_e, $\sigma_n(N_e)$. This procedure is usually started by assuming $\sigma_n = 1/2$ and finding N_e. This N_e is then used to find σ_n which is then modified and a new N_e is found such that the $N_e\sigma_n$ product remains constant for all iterations. This process which is known as the "Debye correction" is described in detail by MATHEWS et al. (1982), Appendix A.

COMBINED TOTAL POWER AND ION LINE POWER SPECTRAL MEASUREMENTS

The so-called ion line feature of the incoherent scatter spectrum yields information which when combined with the corresponding total power measurement not only allows N_e to be better determined but gives two other fundamental plasma parameters. These are the normalized ion neutral collision frequency, ψ_i (to be defined later), and the line-of-sight speed, $v_{\ell s}$ (doppler shift), of the scattering medium. Height variations of ψ_i and $v_{\ell s}$ yield other parameters of aeronomic interest as will be discussed in the next section.

The spectral measurements (backscatter case) are made using pulse-to-pulse techniques. The pulsing scheme is shown schematically in Figure 3. In this scheme in-phase and in-quadrature voltages are measured for each height gate of interest, for each transmitter pulse, and stored individually until needed for
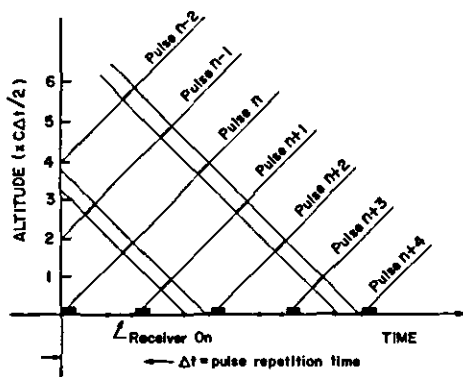
Figure 3. The range-delay diagram appropriate to a "pulse-to-pulse" experiment. Note that any returns from above $C\Delta t/2$ range appear folded together with lower height returns. This is known as range aliasing.

further processing. For our purposes this further processing will be fast Fourier transforming (FFT) some power of two number of complex voltages at each height gate and accumulating the resultant intermediate power spectra. Variations on this signal processing theme (e.g., coherently integrating first) are possible but all essential information is found using the FFT only.

From Figure 3 the interpulse period (IPP) ΔT determines the maximum unaliased bandwidth of the FFT determined power spectrum. This is just $\Delta F = 1/\Delta T$ (e.g., $\Delta T = 1$ msec yields $\Delta F = 1$ KHz) while the number of points N in the FFT determines the frequency resolution $\Delta f = 1/N\Delta T = \Delta F/N$. Each discrete power spectral point given by the FFT is just the average spectral power per unit frequency in that Δf wide region of the spectrum.

The individual power spectral estimates at each height gate are averaged to decrease the variance in the estimate of each spectral point. This approach is known as Bartlett's smoothing procedure for spectral estimators (JENKINS and WATTS, 1968; section 6.3.4). If σ^2 is the power spectral density of an essentially white noise normal process and M is the number of individual spectral estimates which are averaged together then the standard deviation of the estimate of σ^2 at each frequency is $\sigma^2/\sqrt{M}$. Note that increasing the size of the FFT gains only frequency resolution and that each estimate of σ^2 is of the same quality as for "shorter" FFT's. That is, averaging of spectra is necessary and this procedure and the calculation of errors is an exact parallel of what was done for the total power.

In Figure 4 we show a series of normalized and scaled mesospheric spectra obtained at Arecibo (HARPER, 1978). As mentioned in a previous section the power in these ion lines depends on the relative size of the radar wavelength and the Debye length (see Table 4). The ion line widths are seen to decrease with decreasing height and are functions of

$$\psi_i = \frac{\nu_{in}}{\sqrt{2}k_s U_i} \tag{19}$$

where ν_{in} is the ion-neutral collision frequency for momentum transfer, k_s is the scattering wavenumber ($= 4\pi/\lambda_0$), and U_i is the ion thermal speed ($= \sqrt{k_b T_i/m_i}$; m_i = ion mass). As in the case of the electron line, ion line widths can be much narrower than indicated from thermal considerations only. Specifically we find that the collision dominated width (W_{ci}) is very approximately W_{ti}/ψ_i for $\psi_i > 1$ where W_{ti} is the ion thermal spectral width which is about $2\sqrt{2}\ U_i/\lambda_0$. The presence of negative ions tends to increase ion line spectral widths (MATHEWS, 1978; MATHEWS and TANENBAUM, 1981)

MESOSPHERIC SPECTRA
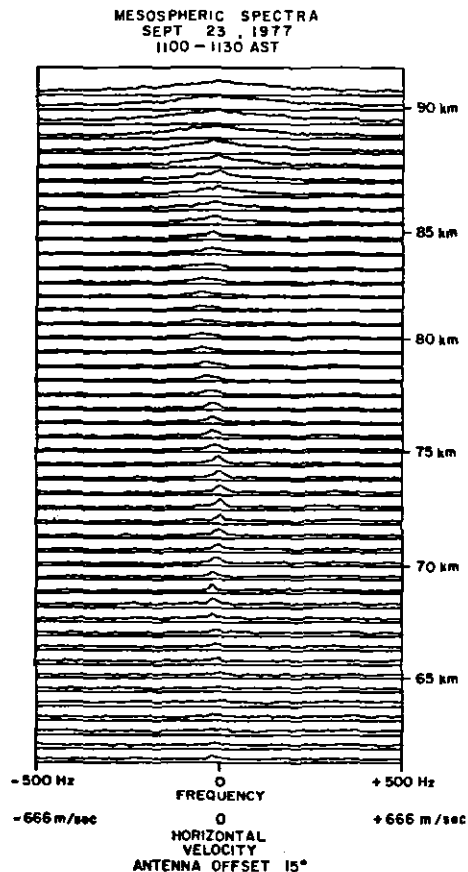SEPT 23, 1977
1100 — 1130 AST

Figure 4. Mesospheric spectra
reported by HARPER (1978).
Note that the spectra become
progressively narrower with
decreasing height.

if electrons and ions are non-reacting chemically on the time scale of interest $(\sim 1/W_{ci})$.

Table 5 contains ψ_i and ion-line widths for the various radars at 70, 80, 90, and 100 km altitude. The collision determined widths (i.e., $\psi_i > 1$) are more appropriate to the 6 dB points of the spectrum. Note that successful application of the pulse-to-pulse ion line spectral technique at a particular altitude requires that $W_i < \Delta F$ (recall $\Delta F = 1/\Delta T$ where ΔT is the IPP) at that altitude. Otherwise aliasing or "folding in" of ion line power from outside the principal band to the principal band occurs.

While it is not clear from Figure 4 each spectrum shown extends from a base power level determined by electron line power and F region power aliased into the principal 1 KHz wide band. The full form of the spectrum is indicated schematically in Figure 5. That power from previous pulses contaminates the wanted signal is clear from Figure 3 where, using a 1 msec IPP as an example, returns from every 150 km range increment are received along with the desired signal. Fortunately the F region spectra have very wide thermal widths and thus, when aliased, appear as the featureless base level shown in Figure 5.

The shape of the spectra shown in Figure 4 are basically Lorentzian. This result can be readily seen from the fluid theory for collision dominated

TABLE 5. APPROXIMATE ψ_i VALUES AND ION LINE SPECTRAL WIDTHS[1]
FOR THE VARIOUS RADARS

Height	ν_{in} (sec^{-1})	EISCAT (224 MHz)		Arecibo/ Millstone $\lambda = 0.69m$		EISCAT (943 MHz)		Sondrestrom	
		ψ_i	W_i(Hz)	ψ_i	W_i(Hz)	ψ_i	W_i(Hz)	ψ_i	W_i(Hz)
100	4.1(3)	0.59	1.12(3)	0.3	2.17(3)	0.14	4.67(3)	0.1	6.46(3)
90	3.2(4)	4.6	243	2.4	900.	1.1	4.2(3)	0.8	6.46(3)
80	1.7(5)	24.	47	12.5	170.	5.8	8.1(2)	4.2	1.54(3)
70	7.7(5)	110.	10	57.	38.	26.	1.8(2)	19.0	3.5(2)

Upper limit of full spectral widths at ∿6 dB down.

Thermal value assuming $T_i = 260$ °K (see text).

Note: Numbers in parentheses are exponents.
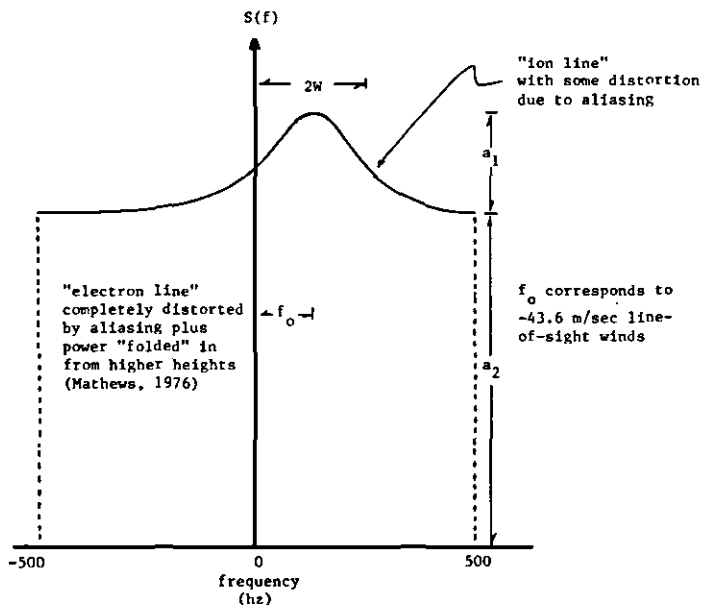
Figure 5. Pictorial representation of an
"ion line" spectrum, similar to those
shown in Figure 4. The spectrum shown
is Lorentzian and thus described by
four parameters. See text.

incoherent scattering in the limit of $\lambda_D \ll \lambda_0$ (TANENBAUM, 1967). In fact in the general case of the collision dominated ion line the shape is also well characterized as a Lorentzian. We therefore recommend non-linear least squares fitting of

$$S(f) = \frac{a_1}{1 + z^2} + a_2$$

(20)

$$z = \frac{f - f_0}{W}$$

to experimental spectra. The four fitted parameters (essentially line width, doppler shift, peak amplitude, and dc level parameters) thus obtained accurately characterize the spectrum while representing a considerable reduction in the amount of data needed for further processing. Also, if the fitted parameters are orthogonal then error per remaining parameter has been reduced by about the square root of the ratio of the number of spectral points to the number of parameters. In practice additional parameters, for example, skewness of the line and higher order denominator terms, might also be obtained. These additional parameters help indicate various interference effects in the manner of SATO and WOODMAN (1982) and could in some cases (narrow lines) be used to correct the power profile as mentioned in the last section.

The effectiveness of the combined total power and ion line power spectral measurements for determining N_e and ψ_i (doppler shift is determined in fitting (20) to the experimental spectra) as functions of height is shown in Figures 6 and 7 for the Arecibo 430 MHz radar and with $T_i = 200°K$. These results have been determined by fitting (20) to a library of theoretical spectra. In Figure 6 the right-hand curve shows how $N_e \sigma_{total}$, which is determined from the radar equations, varies as a function of N_e. The left-hand curves show how the peak ion line power per hertz (times f hertz) varies as a function of N_e for various ψ_i values. The peak ion line power, $N_e \sigma'(f = f_0)$ (x1 Hz), is also found from the radar equation and is related to fitting parameter a_1 in Figure 5. Note that in the previous section we in essence found N_e from the $N_e \sigma_{total}$ curve in Figure 6.

Figure 7 shows the fitted ion line spectral width as a function of N_e for various ψ_i values. Notice then that experimental knowledge of $N_e \sigma_{total}$, $N_e \sigma'(f = f_0)$ (or a_1), and spectral width (W) combined with the three sets of curves yields two separate determinations of N_e and ψ_i if negative ions are absent. If negative ions are present all three sets of curves can be extended into surfaces (see Figure 7 caption) which allow a single determination of N_e, ψ_i, and λ_- where λ_- is the ratio of total negative ion concentration to electron concentration. If the two possible $\lambda_- = 0$ determinations of N_e, ψ_i disagree significantly then N_e, ψ_i, and λ_- should be determined while taking care that both the total power and spectral measurements appear to be free of interference.

SUMMARY OF RECENT RESULTS AND CONCLUSIONS

The use of ISR D-region ion line spectral results to determine line-of-sight winds from the observed doppler shift is straightforward and clearly successful. Results have been reported by MATHEWS (1976), GANGULY (1980), MATHEWS et al. (1981), and FUKUYAMA (1981). The ease of just wind determinations is due to not having to employ the radar equation or details of the scattering mechanism in analysis.

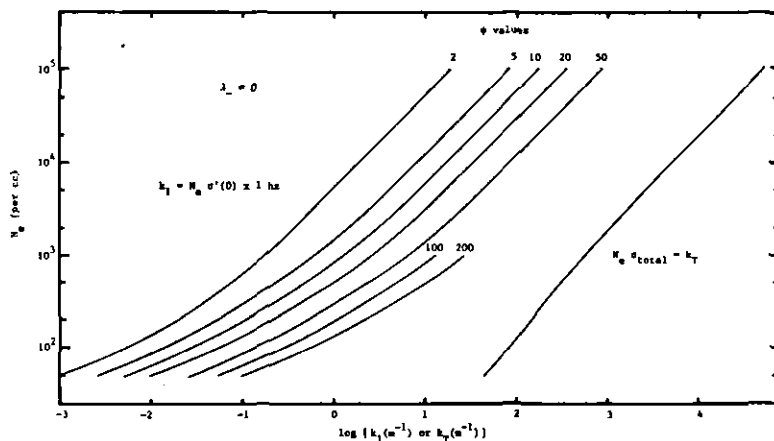D region total power measurements, yielding N_e only, have been reported

Figure 6. Theoretical relationship between
k_T ($=N_e \sigma_{Total}$ where σ_{Total} is the total
scattering cross-section per electron) or
k_1 ($=N_e \sigma'(0)$ is the peak ion line only
scattering cross-section per electron per
hertz) and electron concentration (N_e) for
various ψ values and with $\lambda_- = 0$. The k_1
values are related to spectral parameter
a_1 in equation (20) and Figure 5. Effects
of aliasing have been included in the k_1
versus N_e relationship. Note that combined
total power and spectral measurements plus
the radar equation yield estimates of k_T and
k_1 which for $\lambda_- = 0$ then determined N_e and
ψ information.

recently by TROST (1979) and MATHEWS et al. (1982). The total power measure-
ments are easy to do and it is a relatively easy and accurate (subject to some
of the caveats raised in the preceding section) process to convert total power
measurements to electron concentration. Measurements yielding N_e only are of
course quite useful for investigation of D-region chemical/production processes.
For example, at mid and low latitudes the electron concentration, which is a
function of height and solar zenith angle, can be used in conjunction with
chemical models to study some of the major processes in D region photochemistry.
At high latitudes N_e determinations can be used to directly study the
energetic particle mechanism for production of the ionospheric plasma.

The combined total power and ion line spectral measurements of course yield
more than N_e information about the mesosphere. This fact is made admirably
clear by Figure 7 and 8 of FUKUYAMA (1981). These figures, shown below as
Figures 8 and 9, give sequences of ion line half width and power measurements
made in Arecibo. These sequences clearly show wavelike features which require
further interpretation. The required interpretation is probably best done in
terms of N_e and ψ_1 which can be found from combined total power and spectral
measurements in the manner described in the previous section.

The N_e and ψ_1 values determined in this manner depend on having assumed
ion temperatures and mean masses (using, for example, CIRA temperatures and 31
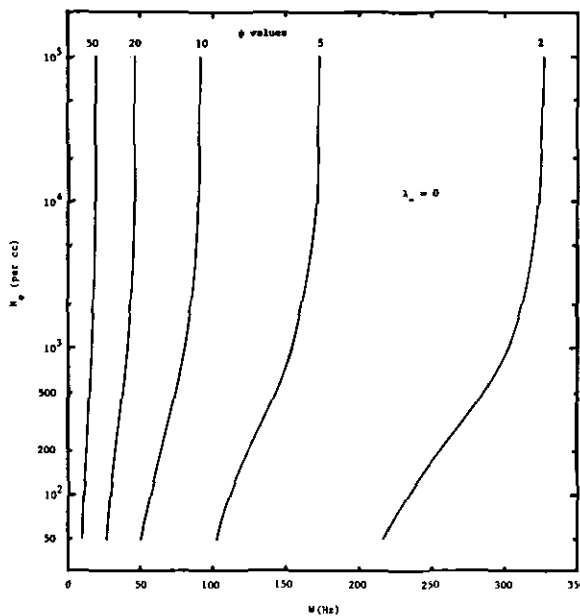amu ions) but these "intermediate" values can be converted to absolute values.

Figure 7. Theoretical relationships between spatial
width (W) and N_e for various ψ values and $\lambda_- = 0$.
These results assume a Lorentzian ion line spectral
shape but with aliasing included. Note that for
$\lambda_- = 0$ Figure 6 together with these results over-
determine one parameter. If $\lambda_- \neq 0$ then all curves
in Figure 6 and above become surfaces with λ_- lying
along the third axis (out of the page).

This is done by assuming that over a small height region the whole variation of
ψ_i with height is due to changing ion neutral collision frequency which is
pressure dependent. We thus let

$$\nu_{in} = \nu_{in}(z_0) \exp\left[-\frac{(z-z_0)}{H}\right] \qquad (21)$$

and find H, the local scale height, from the slope of the $\ln \psi_i$ variation
(TEPLEY and MATHEWS, 1978). The T_i/m_i ratio is then used to get the
absolute ψ_i values and thus also ν_{in} and pressure values. The value of
N_e will change typically by less than 5% since only the Debye correction
changes. Results using these methods have recently been reported by TEPLEY et
al. (1981).

If scattering from "chemical fluctuations", a possibility reported by
KOCKARTS and WISEMBERG (1981) and WISEMBERG and KOCKARTS (1983), can be ignored
then negative ion concentrations can be found as mentioned in the previous
section. Results obtained in this manner have been reported by GANGULY et al.
(1979), however, further ISR aeronomic work concerning negative ions should
await the sorting out of the theoretical problems.

In conclusion we find that all the current incoherent backscatter radars
can make D-region measurements at least under some conditions. These radars
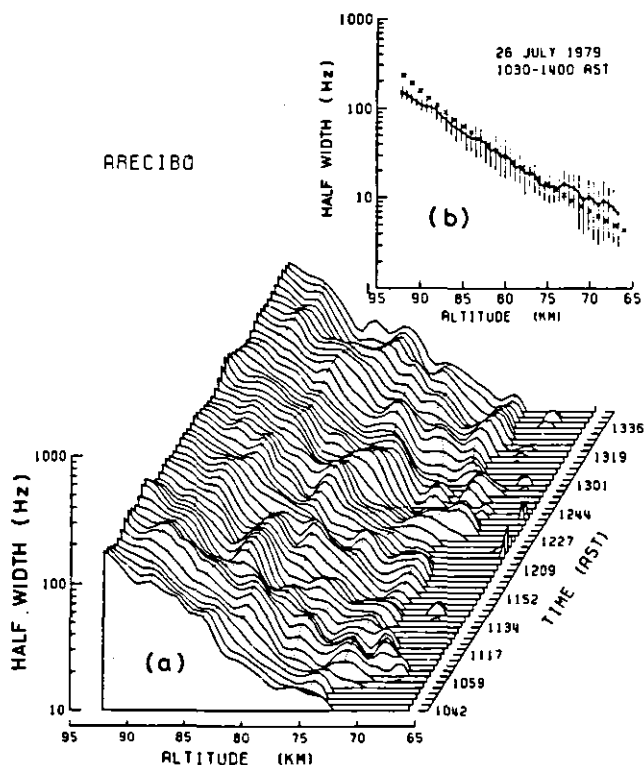have been discussed in previous sections in terms of their practical

Figure 8. Ion line half widths versus height and time
(a) or time averaged (b). The time averaged results
(b) are compared with widths calculated using CIRA
1972. Note the considerable wave structure seen
(FUKUYOMA, 1981).

capabilities and limitations. For example, both the EISCAT 934 MHz facility and
the Sondrestrom facility require the high electron concentrations characteristic
of auroral activity in order to "see" into the lower D region.

The bistatic radars have not been discussed here because of the
complexities, due to overlapping antenna patterns, associated with absolute
calibration. These radars can however measure spectral doppler shifts with the
same limitations as the radars discussed.

Future uses of incoherent scatter radars in the D region should involve
multiple radars, for example, Arecibo and Millstone Hill. Hopefully other
techniques such as rocket-borne mass spectrometers and electron concentration
sensing probes will be flown near to, say, Arecibo in an attempt to cross
compare and verify the various techniques and perhaps even shed light on the
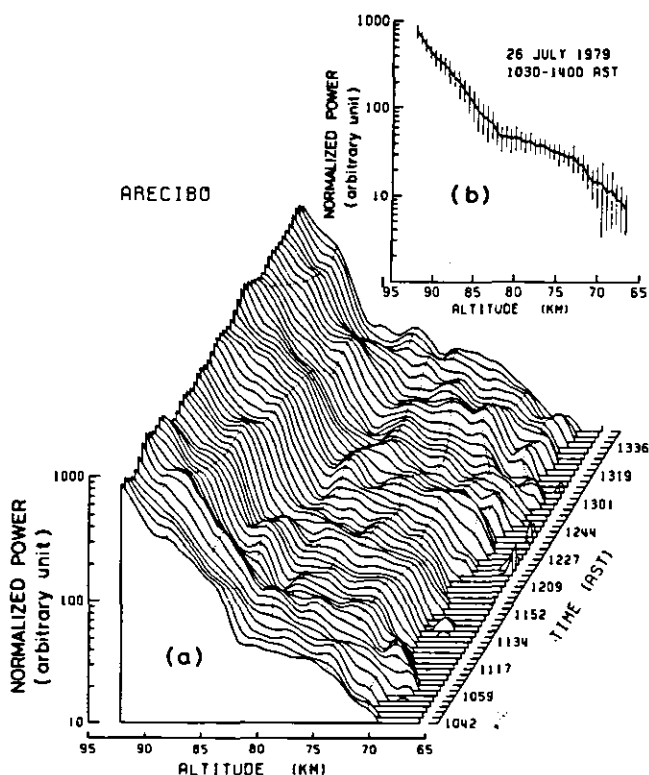lower D-region scattering mechanism.

ACKNOWLEDGEMENT

Figure 9. Similar to Figure 8 but for ion line power (FUKUYAMA, 1981).

REFERENCES

Armistead, G. W., J. V. Evans and W. A. Reid (1972), Measurements of D- and E-region electron densities by the incoherent scatter technique at Millstone Hill, Radio Sci., 7, 153-162.

Bekefi, G. (1966), Radiation Processes in Plasmas, chapter 8, John Wiley & Sons, Inc., New York.

Berne, B. J. and R. Pecora (1976), Dynamic Light Scattering, chapter 3, John Wiley & Sons, Inc., New York.

Breakall, J. K. and J. D. Mathews (1982), A theoretical and experimental investigation of antenna near-field effects as applied to incoherent back-scatter measurements at Arecibo, J. Atmos. Terr. Phys., 44, 449-454.

Dougherty, J. P. and D. T. Farley (1960), A theory of incoherent scattering of radio waves by a plasma, Proc. Royal Soc. A, 259, 79-99.

Dougherty, J. P. and D. T. Farley (1963), A theory of incoherent scattering of radio waves by a plasma, 3. Scattering in a partially ionized gas, J. Geophys. Res., 68, 5473-5486.

Evans, J. V. (1969), Theory and practice of ionospheric study by Thomson scatter radar, Proc. IEEE, 57, 496-530.

Farley, D. T. (1966), A theory of incoherent scattering of radio waves by a plasma, 4. The effect of unequal ion and electron temperatures, J. Geophys. Res., 71, 4091-4098.

Farley, D. T. (1969), Incoherent scatter power measurements; a comparison of various techniques, Radio Sci., 4, 139-142.

Feynman, R. P., R. B. Leighton and M. Sands (1965), Lectures on Physics, Vol. I, section 34-9, Addison-Wesley, Reading, Mass.

Fukuyama, K. (1981), Incoherent scatter radar observations of wavelike structures in the mesosphere over Arecibo, J. Geophys. Res., 86, 9152-9158.

Ganguly, S. (1980), Incoherent scatter observations of mesospheric dynamics at Arecibo, Geophys. Res. Lett., 7, 369-372.

Ganguly, S, J. D. Mathews and C. A. Tepley (1979), Thomson scatter radar detection of D-region negative ions at Arecibo, Geophys. Res. Lett., 6, 89-92.

Gray, R. W. and D. T. Farley (1973), Theory of incoherent scatter measurements using compressed pulses, Radio Sci., 8, 123-131.

Harper, R. M. (1978), Preliminary measurements of the ion component of the incoherent scatter spectrum in the 60-90 km region over Arecibo, Geophys. Res. Lett., 5, 784-786.

Ioannidis, G. and D. T. Farley (1972), Incoherent scatter observations at Arecibo using compressed pulses, Radio Sci., 7, 763-766.

Ioannidis, G. and D. T. Farley (1974), High resolution D-region measurements at Arecibo, Radio Sci., 9, 151-157.

Jenkins, G. M. and D. G. Watts (1968), Spectral Analysis and its Applications, Holden-Day, San Francisco.

Kittel, C. (1958), Elementary Statistical Physics, chapter 25, John Wiley, New York.

Kockarts, G. and J. Wisemberg (1981), Chemical fluctuations and incoherent scattering theory in the terrestrial D region, J. Geophys. Res., 86, 5793-5800.

Kraus, J. D. (1950), Antennas, McGraw-Hill, New York.

Lalonde, L. M. (1966), Incoherent backscatter measurements of sporadic E, J. Geophys. Res., 71, 5059-5065.

Mathews, J. D. (1976), Measurements of the diurnal tides in the 80- to 100-km altitude range at Arecibo, J. Geophys. Res., 81, 4671-4677.

Mathews, J. D. (1978), The effect of negative ions on collision-dominated Thomson scattering, J. Geophys. Res., 83, 505-512.

Mathews, J. D., J. K. Breakall and S. Ganguly (1982), The measurement of diurnal variations of electron concentration in the 60-100 km ionosphere at Arecibo, J. Atmos. Terr. Phys., 44, 441-448.

Mathews, J. D., M. P. Sulzer, C. A. Tepley, R. Bernard, J. L. Fellous, M. Glass, M. Massebeuf, S. Ganguly, R. M. Harper, R. A. Behnke and J. C. G. Walker (1981), A comparison between Thomson scatter and meteor wind measurements in the 65-105 km altitude region at Arecibo, Planet. Space Sci., 29, 341-348.

Mathews, J. D. and B. S. Tanenbaum (1981), A plasma wave and electron-plasma diffusion interpretation of Thomson scattering from a plasma containing negative ions, Planet. Space Sci., 29, 335-340.

Reitz, J. R. and F. J. Milford (1967), Foundations of Electromagnetic Theory, Addison-Wesley, Reading, Mass.

Rihaczek, A. W. (1969), Principles of High Resolution Radar, pp. 159-225, McGraw-Hill, New York.

Sato, T. and R. F. Woodman (1982), Spectral parameter estimation of CAT radar echoes in the presence of fading clutter, Radio Sci., 17, 817-826.

Sechrist, Jr., C. F. (1974),. Comparisons of techniques for measurement of D-region electron densities, Radio Sci., 9, 137-149.

Stutzman, W. L. and G. A. Thiele (1981), Antenna Theory and Design, John Wiley & Sons, Inc., New York.

Taylor, G. N. (1975), Incoherent scatter measurements of E- and upper D-region ionization during three solar flares, J. Atmos. Terr. Phys., 37, 349-357

Tepley, C. A. and J. D. Mathews (1978), Preliminary measurements of ion-neutral collision frequencies and mean temperatures in the Arecibo 80-100-km

154

altitude region, J. Geophys. Res., 83, 3299-3302.

Tepley, C. A., J. D. Mathews and S. Ganguly (1981), Incoherent scatter radar studies of mesospheric temperatures and collision frequencies at Arecibo, J. Geophys. Res., 86, 11330-11334.

Thome, G. D. and L. S. Wagner (1971), Electron density enhancements in the E and F regions of the ionosphere during solar flares, J. Geophys. Res., 76, 6883-6895.

Trost, T. F. (1979), Electron concentrations in the E and upper D region at Arecibo, J. Geophys. Res., 84, 2736-2742.

Wisemberg, J. and G. Kockarts (1983), A new ionospheric scattering mechanism, J. Atmos. Terr. Phys., 45, 47-53.

UNPUBLISHED MATERIAL

Breakall, J. K. (1983), On the absolute calibration and theoretical justification of high resolution ionospheric results obtained from Arecibo radar measurements, Ph.D. Thesis, Case Western Reserve University, Cleveland, Ohio

Lalonde, L. M. (1967), D and E region studies at Arecibo, in Thomson Scatter Studies of the Ionosphere - An Informal Conference Record, Aeron. Rep. No. 19, Aeronomy Laboratory, University of Illinois.

12. BINARY PULSE COMPRESSION TECHNIQUES FOR MST RADARS

R. F. Woodman*, M. P. Sulzer** and D. T. Farley***

*Instituto Geofisico de Peru
Lima, Peru

**Arecibo Observatory, Arecibo, Puerto Rico   00613

***Cornell University, Ithaca, NY   14853

INTRODUCTION

In most MST applications pulsed radars are peak power limited and have excess average power capability. Short pulses are required for good range resolution, but the problem of range ambiguity (signals received simultaneously from more than one altitude) sets a minimum limit on the interpulse period (IPP). Pulse compression is a technique which allows more of the transmitter average power capacity to be used without sacrificing range resolution. As the name implies, a pulse of power P and duration T is in a certain sense converted into one of power nP and duration T/n. In the frequency domain, compression involves manipulating the phases of the different frequency components of the pulse. A short pulse consists of contributions from a wide band of frequencies, all of which are in phase at one point in space-time. Changing the phase relations on transmission lengthens the pulse, but it can be reassembled into a short pulse upon reception by proper processing <u>if</u> the phase has not been perturbed in some unknown way in the meantime (i.e., by the scattering process). This is essentially the idea behind frequency "chirping".

Another way to compress a pulse is via phase coding, especially binary phase coding, a technique which is particularly amenable to digital processing techniques. This method has been used extensively in recent years in radar probing of the atmosphere and ionosphere, and it is the method we will discuss here. The general topic of pulse compression is dealt with in COOK and BERNFELD (1967), BARTON (1975), BROOKNER (1977), and other texts. The operation performed on reception to produce the compression of phase coded radar signals is referred to as decoding. The signal is decoded upon reception by passing it through a "filter" whose impulse response is the reverse in time of the transmitted pulse (the pulse "played backwards", so to speak). Such a filter is said to be "matched" to the pulse. In practice these matched filters are usually specially designed acoustic surface wave devices or conventional filters plus digitizers, digital delay lines, and some add/subtract circuitry or equivalent software.

From another point of view, the decoding process consists of cross-correlating the received signal with a replica of the transmitted pulse; hence, if an echo from a single scattering target is passed through such a decoder, the output is the autocorrelation function (ACF) of the pulse delayed by a time corresponding to the distance to the target. As an example, the phase coding sequence and the ACF of a 5-baud Barker coded pulse are listed below.

$$+ \ + \ + \ - \ +$$

$$. . . 0 \ 0 \ 0 \ 1 \ 0 \ 1 \ 0 \ 5 \ 0 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0 . . .$$

If the compression process were perfect, only the 5 would be present in the above ACF; the 1's represent undesired range sidelobes.

156

In this paper we discuss many aspects of codes and decoding and their applications to MST experiments. This includes Barker codes and longer individual codes, and then complementary codes and other code sets. We discuss software decoding and also hardware decoders and coherent integrators.

BARKER CODES

A class of codes known as Barker codes (BARKER, 1953) has been used extensively in ionospheric incoherent scatter measurements. The Barker coded pulse is considered to be made up of n "bauds", each of duration T, so the total duration is nT, with the maximum value of n being 13. The phase of each baud is 0 to 180 degrees (± 1), in a sequence that depends on n.

We have seen above the properties of a Barker code of length 5. A property of all Barker codes is that the center peak of the correlation function is n, and the sidelobes are either 0 or 1. For ionospheric application the sidelobes are generally not a problem since, for n equal 13, say, the power corresponding to the central peak is 169 times greater than that in each of the 12 sidelobes. (Note that the signal-to-noise ratio in the central peak is increased by the compression by a factor of 13, not 169, since the noise is the sum of 13 independent samples.)

The above discussion is valid for scatter probing of the atmosphere as long as the correlation time of the scattering medium is long compared to the total (uncompressed) duration of the coded pulse. In practice this is always the case of MST applications but may not be true for incoherent scatter from the ionosphere, for example. Detailed calculations of what happens in the latter case are given by GRAY and FARLEY (1973), and a general discussion of the "ambiguity function" of a Barker coded pulse as a function of target induced Doppler shift is given in COOK and BERNFELD (1967). Gray and Farley also discuss the use of multiple coded pulse sequences in the measurement of the ACF of the scattering medium. The effect of the coding is usually minimal; in typical situations the "true" ACF is convolved with a function whose width is about one baud. Finally, although 13 baud is the longest possible Barker sequency (unity sidelobes), there are many longer sequences with sidelobes that are only slightly larger. As an example, a 28-baud sequence with a maximum sidelobes level in the ACF of 2 is listed by Gray and Farley. In fact, this code sequence was used in the first application of phase coding techniques in MST radars (Woodman, MPI-SOUSY (SCHMIDT et al., 1979) and the Arecibo 430 MHz radar (WOODMAN, 1980a).

Complementary codes are again binary phase codes and they come in pairs. They are decoded exactly as are Barker codes, by a "matched" filter/delay line combination whose impulse response is the time reverse of the pulse. The range sidelobes of the resulting ACF output will generally be larger than for a Barker code of comparable length, but the two pulses in the complementary pair have the property that their sidelobes are equal in magnitude but opposite in sign, so that when the outputs are added the sidelobes exactly cancel, leaving only the central peak; i.e., the compression is perfect. As the simplest possible example, consider the 2-baud complementary pair below.

```
Code:              + +        (first pulse)
                   + -        (second pulse)

ACF:   0 +1 +2 +1 0          (first pulse)
       0 -1 +2 -1 0          (second pulse)
       0  0 +4  0 0          (sum)
```

Representing the above pair as (A,B) it is easy to show that the sequence (AB, A$\overline{B}$), where $\overline{B}$ is the complement of B, is also a complementary pair that is twice as long. Proceeding in this way one can obviously generate long n-baud code

pairs, where n is any power of two. It turns out that n can also be ten, or ten times any power of two. Further properties of these sequences are given by GOLAY (1961). In the first reported measurements (SCHMIDT et al., 1979; WOODMAN, 1980) n was 32, compressing a 32 μsec pulse to 1 μs to achieve an altitude resolution of 150 m.

There are two practical limitations on the maximum value of the compression ratio n: (1) as n increases the effect of ground clutter extends to higher and higher altitudes; (2) the computing requirements for decoding increase with n. The first is the most serious limitation; the computing requirements usually can be handled one way or another. One process that often simplifies the computing is coherent integration (summing N successive voltage samples from a given altitude before doing any other processing). Since coherent integration and decoding are linear operations they can be interchanged (WOODMAN, 1980a); e.g., samples from 100 pulses, say, can be coherently integrated and then decoded all at once. In dealing with the first limitation one must achieve some compromise between three competing goals: (1) the desire to confine strong ground clutter effects to the lowest possible range of altitudes (i.e., use short pulses); (2) the desire to avoid range ambiguity (use a long IPP); and (3) the desire to use the full average power capabilities of the transmitter to achieve maximum sensitivity.

MORE COMPLEX COMPLEMENTARY CODING SCHEMES

WOODMAN (private communication, 1980) suggested that it might be useful to transmit more complex complementary sets in order to partly alleviate the ground clutter/range ambiguity problem. The cross correlation function (XCF) of the basic complementary transmitted sequence, A, B, A, B, ... with the decoding function A, B is periodic with a period 2T, where T is the interpulse period (between A and B), but there are also substantial nonzero values of the XCF in the vicinity of T. For example, for the 4-baud pair (++-- +-+) the XCF is

$$\ldots\ 0008000\ \ldots\ 0040400\ \ldots\ 0008000\ \ldots$$

At delays near T from the "wanted" return, the range sidelobes of the individual pulses add rather than cancel, whereas the main peak does cancel. The 4's in the above represent the most important source of range ambiguity. These can be eliminated by transmitting the more complex sequence, A, B, A, B̄, A, B, A, B̄, ... and decoding by cross correlating with A, B, A, B̄. The XCF for this scheme consists of single identical spikes at intervals of 2T; i.e., the first range sidelobe is pushed out to twice the interpulse spacing. By extending this idea the first sidelobe can be pushed out to even higher multiples of T. In this way a substantial range of altitude could be probed at a very high pulse repetition frequency (PRF). In actual practice, though, some altitudes would be lost because of the necessity of blanking the receiver during actual pulse trans- mission and because of receiver saturation by ground clutter. GONZALES and WOODMAN (1981) have used long sequences of complementary codes for a partial reflection experiment using the Arecibo HF-Heating facility. They were able this way to avoid multiple reflections from F and E region heights.

QUASI-COMPLEMENTARY CODE SETS

The results presented so far have all been based on the assumption that the transmitted pulses were perfectly coded. In practice of course this won't be true; the phase shifts will require a finite amount of time and will not be exactly 180 degrees, etc. As a result the range sidelobes for the complementary code pairs will not cancel exactly; the location of the sidelobes will depend on what sort of error is made by the transmitter. SULZER and WOODMAN (1983) have developed a technique to minimize this problem. Rather than transmit just a pair of complementary 32-baud codes, they transmit a sequence of 48 different

32-baud pulses. Each is decoded individually and the results are combined
coherently, so in a sense the whole sequence can be considered to be a single
code. But from another point of view we can think of the sequence as 24 quasi-
complementary pairs, each with a different set of small range sidelobes, due
partly to errors in transmission and partly to the fact that the pairs are not
perfectly complementary. Because the sidelobes produced by the individual pairs
have a more or less random distribution, the resultant sidelobes of the entire
sequence are lower and more uniform than those of a single (imperfect) comple-
mentary pair.   This is no accident of course; the codes were chosen by an
extensive computer search requiring about 350 hours (1) using a Harris computer
and an FPS AP120B array processor. The major disadvantage of this technique is
that no coherent integration before decoding is possible; at present only the
Arecibo Observatory has the digital preprocessing equipment required for the
extensive high speed decoding.

A similar idea has been developed by the same authors for mesospheric
observations at Arecibo. To achieve the desired resolution of 600 m (4 µs) and
fully utilize the transmitter, one would ideally use a 52-baud Barker code,
which unfortunately does not exist and 4 msec IPP to avoid range fading of F-
region scatter. A good approximation to this can be achieved by a pseudorandom
sequence of pseudorandom 52-baud codes found by a 10 hour computer/array
processor search. Sidelobe power is not reduced to the level of a Barker code,
but its coherence from pulse to pulse is destroyed by use of a different code on
each transmitted pulse. The only range retaining the spectral information of
the medium is the main lobe.

PSEUDORANDOM PERIODIC CODES

In bistatic experiments, one can transmit continuously, but one wishes to
code and decode in such a manner that it is equivalent to transmitting a
sequence of short pulses. Codes used for this are the so called maximal –
length pseudorandom sequences which can be generated by relatively simple con-
figurations of flip-flops. These are excellent codes, having a uniform sidelobe
level of -1 and lengths of any power of two minus one.

DECODING

As mentioned in the introduction decoding can be considered either as a
matched filter operation, i.e., a convolution, with a filter which has an input
response equal to the pulse shape, or as a cross correlation of the input with
the pulse shape. We can perform this operation by proper software in a general
purpose computer, in a special purpose processor or in an analog device
(filter).   Here we shall describe different techniques which have been used for
this purpose as well as some new ideas that have been proposed with special
advantage for MST radar applications.

SOFTWARE DECODING

Decoding, if done in a straight forward manner, would normally take several
tens of operations per µsec and would be out of the possibilities of moderately
priced computers. Decoding is used to obtain optimum resolution, which in the
case of MST radars is of the order of a few hundred meters, this corresponds to
one complex sample per one or few µsec. If we take a 32 baud code with a band-
width of 1 µsec, as an example, straight decoding would take 64 adding or sub-
tracting operations per µsec; certainly a requirement beyond the capabilities
of even the fastest computers and a very demanding one even for specially built
digital equipment. Fortunately, MST radars at VHF and lower UHF frequencies
produce highly redundant information. Correlation times in the medium, and
hence of the echo signals they produce, are of the order of a fraction of a
second (=0.5 to 2 sec for 50 MHz). This calls for coherent integration from

pulse to pulse, when performed, reduces the amount of information by as much as two orders of magnitude.

Decoding and coherent integration are linear operations and, as pointed out by WOODMAN et al. (1980), they commute. One can perform the coherent integration first, and decode afterwards, with identical results. This possibility permits performing decoding operations in just about any mini- or micro-computer available in the market.

Coherent integration is so efficient in reducing the amount of information, that this can be put in a few tapes per hour of observation. In this case *decoding, statistical processing and parameter estimation can then be performed off-line.* This approach has been taken for the SOUSY radar - using a hardware integrator - (SCHMIDT et al., 1979) and by the same group at Arecibo with a portable 50 MHz transmitter.

Coherent integration usually requires a special purpose digital equipment, but can be done with fast array processors when used at the front end of the processing system. The M-mode at Millstone Hill uses this approach using an AP-120B for the integration and decoding (RASTOGI, 1983 private communication).

Jicamarca, at present, performs low resolution decoding off-line, by coherently integrating on line with a Harris/6 general purpose computer.

HARDWARE DECODERS: A COHERENT INTEGRATOR-DECODER

Coherent integration by general purpose computers usually takes most of the computer power to perform this task, leaving no CPU time to perform the decoding and statistical processing. Therefore, it is highly recommended to perform the coherent integration in a dedicated device. A device constructed to do this task can perform the decoding operation with very little added complexity.

A coherent integrator and decoder has been designed and built for the SOUSY radar (WOODMAN et al., 1980). To the authors' knowledge, it is the only device specially designed for MST applications which has been built for this purpose. The device is described in some detail in the cited reference. Here we shall limit ourselves to reproducing the block diagram (Figure 1) and discussing some of its features.

The W.K.R. decoder is a programmable device. It also performs the function of a radar controller. All the parameters which control the radar and data taking sequence are programmed in 4 computer addressable PROMS. The parameters include Transmitter IPP, baud length, code length, sequence, sampling rate and groupings. For this purpose it interprets 16 program instructions which are accompanied by a number that specifies a time delay before next instruction is executed. Maximum sampling rate is 0.5 μsec, it can process 1024 altitudes (before decoding) with single codes, or 512 in the case of code pairs like complementary codes.

THE ARECIBO PLANETARY DECODER - A CROSS-CORRELATOR

Apart of the W.K.R. decoder, the only other hardware decoder that has been used for a MST radar is the Arecibo Planetary Decoder (schematics and some text in a maintenance document at the Arecibo Observatory). The device was built for planetary radar, a more demanding application. It is capable of performing 1000 additions or subtractions per μsec and has a maximum sampling rate of 1 μsec. It has been used for decoding at the front end of any previous processing, since it has sufficient computing power to perform decoding (as much as 256 x 1 μsec bauds) without any previous coherent integration.
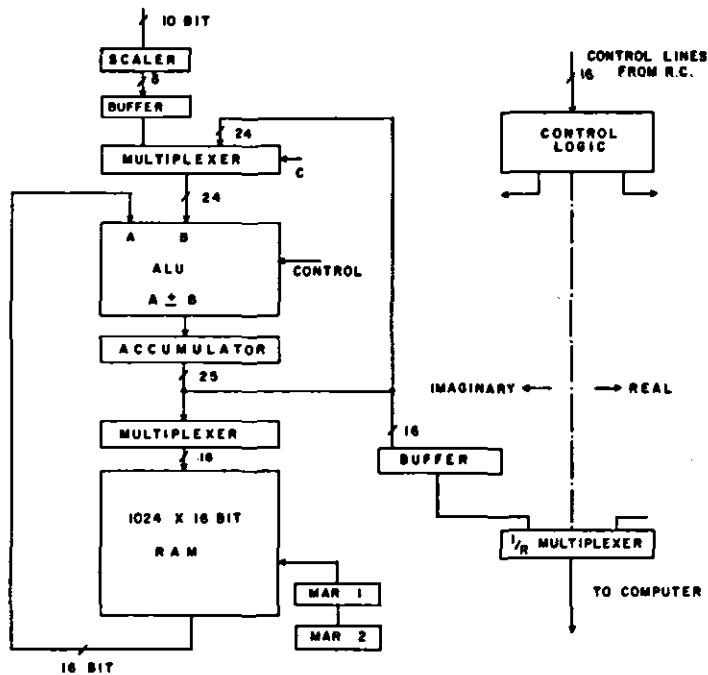
Figure 1. Block diagram for WKR Decoder and
coherent integrator for SOUSY radar.

The device is more powerful than required by MST applications, but it was
already available. It has been used to decode and coherently integrate (in that
order) 32 baud complementary code sequences in conjunction with the 430 MHz
radar (WOODMAN, 1980). Its speed has allowed the implementation of quasi-
complementary sequences. These are long sequences of transmitter pulses, each
coded with a different code which present advantages that have been discussed by
SULZER and WOODMAN (1983) but which do not permit coherent integration before
coding.

The Arecibo Planetary Decoder consists, essentially, of four parallel cross
correlators. Each cross correlator consists of a selectable number 2048 maximum
of lag-product integrators, depicted schematically in Figure 2. Each integrator
register is associated with a given lag, and integrates the product of the
present, 10 bit digitized analog signal, "multiplied" by the delayed value of a
2-bit (3 level, also called 1 1/2 bit) sample of the other signal (code in our
case). The delay for the delayed signal, is produced by a 2-bit shift register.
Each integrator is associated with one of the registers of the shift register
and, therefore, to a given lag delay. The necessary multiplications and
updating additions are performed by 64 parallel adders per cross-correlator.
Speed is achieved by parallel operations and by the fact that multiplying a
3-level signal involves only additions, subtractions and no-operations (1,-1,0).

When a correlator is used as a decoder as shown in Figure 2, the length of
the code is in principle unlimited, although the accumulator must eventually
overflow. However, the number of ranges is limited to the number of lags. If
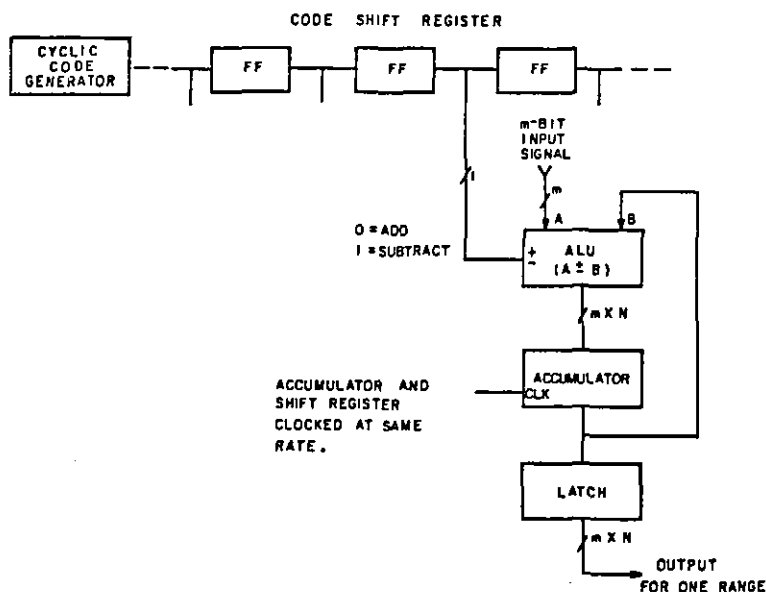the code length is equal to or less than the number of lags, the number of

Figure 2. Decoder using correlator architecture.

ranges is also unlimited since each accumulator can be dumped as it completes the last bit of the code, and the code can be fed to the correlator again so that the lags are used for another set of ranges.

DECODING PERIODIC PSEUDORANDOM CODES

Decoding of bi-static radars, where the atmosphere is illuminated by a continuous but coded wave, requires special discussion. The length of the code in this case is much larger than in the case of a pulsed radar. It has to be at least as long as the range of altitudes one expects to receive echoes, to prevent range folding of the echoes. This can put demands on the decoding operation which are orders of magnitude higher than in pulsed MST radars.

So far, the only ST radar which works in a bi-static coded mode is the Arecibo 2380 MHz radar (WOODMAN, 1980b). In this case pseudorandom phase coded sequences with a baud length of 0.2 µsec and a period of 1024 baud where used. This corresponds to a resolution of 30 meter with a range ambiguity of ≈30 km. Decoding at all ranges would involve ≈500 operations per µsec, a formidable requirement, even when we reduce the number of ranges. Coherent integration does not help much in this case since the coherence time at 2380 MHz is nearly two orders of magnitude shorter than at 50 MHz. Nevertheless, decoding of these signals was accomplished by means of the Arecibo 1008 channel correlator (HAGEN, 1972).

Decoding becomes possible, because of the time stationarity of the radar returns, since the codes have constant amplitude and are transmitted continuously. This time stationarity permits the use of one-bit and three level (1 1/2 bit) cross correlators, which can be economically implemented by parallel integration with a large number of lags at a high sampling rate. The Arecibo correlator has 1008 lags and a maximum sampling rate of 20 MHz. This represents 20,000 operations per sec, which in this case involves simply the counting up or down of 1024 parallel counters.

The correlator at Arecibo was built for the spectral analysis of broadband (10 MHz) radio astronomical signals, by evaluating its autocorrelation function. Its configuration is similar to the one depicted in Figure 3. The main difference being that the multiplication and addition are replaced by a 2 bit multiplier and that accumulation is performed by a counter. The correlator is divided into four identical units of 252 lags each. The scheme reported by Woodman used two of them in parallel, one for the real and one for the imaginary, with 252 lags on each. This limits the range of observations to 7 1/2 kms, which is more than enough, since the range of altitudes of interest is from ≈14 to 19 kms. The radar cannot illuminate below 14 kms and there is no sensitivity beyond 19 kms.

The Arecibo correlator has one limitation: it takes too long to dump the information. It takes so long that two consecutive decoded profiles do not have any correlation in between. This lack of correlation prevented the determination of either the velocity or spectral width of the echoes. This limitation has been circumvented recently by using the other half of the correlator in what corresponds to a double-buffered scheme (WOODMAN and IERKIC, 1983 private communication).

At present a new correlator is being built at Arecibo which will have a buffered dump, with practically no dumping delay between decoded profiles. This should allow the evaluation of full spectral information in the future.

TRANSVERSAL FILTERS USED AS DECODERS

A decoder performs a convolution

$$o(t) = i(t) * h_j(t) \tag{1}$$

where $i(t)$ is the input signal
$h_j(t)$ is the impulse response of the decoder programmed for code j
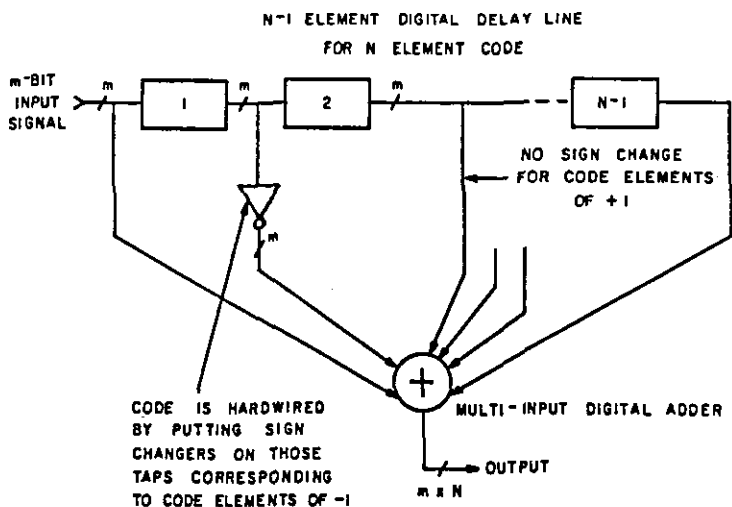
and $o(t)$ is the output signal.

Figure 3. Transversal filter decoder for code length N.

The convolution may be written as an integral

$$o(t)=\int_{-\infty}^{t} i(q) \ h(t-q)dq \tag{2}$$

This integral becomes a summation for discrete samples. Figure 3 shows a block diagram of a decoder for a code N long. The present value and N-1 earlier values are available in a delay line. In the case of a binary phase code the impulse response consists of +1 or -1 and so the multiplications are replaced with additions of either the signal or the inverted signal. Figure 3 shows a digital implementation. This could be done with an analog line and analog inverter and adders.

If we wish to change the code quickly, we must add a shift register and selector switches. This is shown in Figure 4. Note that if the signal enters from the left, the code must fill the shift register from the right. This accomplishes the folding about the time axis which is indicated by the negative sign in equation 2.

For a practical decoder, we need one more shift register which is parallel to the one storing the phase code. This is the amplitude of the code and is "1" wherever the code is defined and "0" otherwise. It is also attached to the control lines on the switches and when it is "0" neither switch is connected. This makes it possible to use codes with the length less than N.

An example of the type of decoder described in Figure 3 is the Barker Decoder used at Arecibo Observatory. The device is completely analog and works at the 30 MHz I.F. frequency. The input signal is fed to a surface-wave delay line with taps. The appropriate taps are inverted and added. The device is good only for the 13-bit Barker code with 4 µsec baud.

The transversal filter decoder would probably not be built with a digital implementation because of the complexity of the multi-bit delay line, and the adder. With the correlator design described in the last section, only the 3 level signal need be delayed, and one can choose either unlimited code length or range. For digital implementation, the correlator configuration is clearly the best.
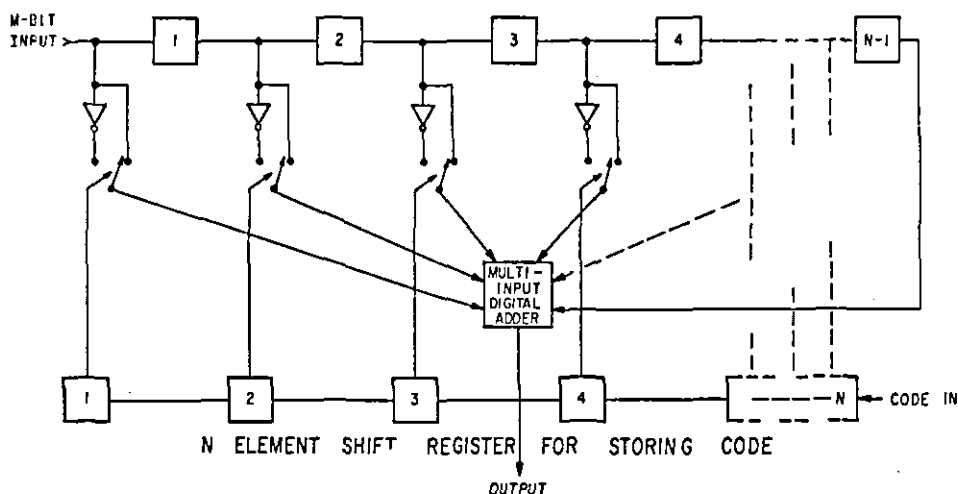

Figure 4. Transversal filter decoder for code
length N (programmable code).

NEW IDEAS FOR NEW DECODERS: AN ANALOG RING DECODER

When designing a filter which will have an analog input and output, we must consider the possibility that using analog circuitry inside may be the best approach. High speed digital processors are very powerful machines, but they have some disadvantages. The most important disadvantage for our purpose is difficulty in maintenance. These machines are usually one, or a few, of a kind and the trouble-shooting procedures are often not well-defined.

We present here a new analog design which is similar to the normal transversal filter design in that the output is the simultaneous sum of N stored samples but is like the correlator design in that only the code, a simple three-level signal, need be shifted. This design has been proposed by SULZER and WOODMAN (1983, MAP Conference).

Figure 5 is a block diagram showing the concept of the analog ring decoder. To understand its operation first look at the rotor of the central rotary switch which is connected to the input signal. This switch deposits samples on the capacitors which are stored until the switch completes a revolution and deposits a new sample. As long as the code is shorter than the number of capacitors, we have enough information stored to decode the signal. All we have to do is add the signals from N adjacent capacitors with the correct polarity. To do this, we place the code in a shift register, and the outputs of the shift register control switches in the amplifiers to give the proper gain (+1, -1, or 0) to the signal. Everytime the switch moves forward one step, so does the shift register, and the correct signals are given to the delayed signals, which are added through the resistors to the output bus.
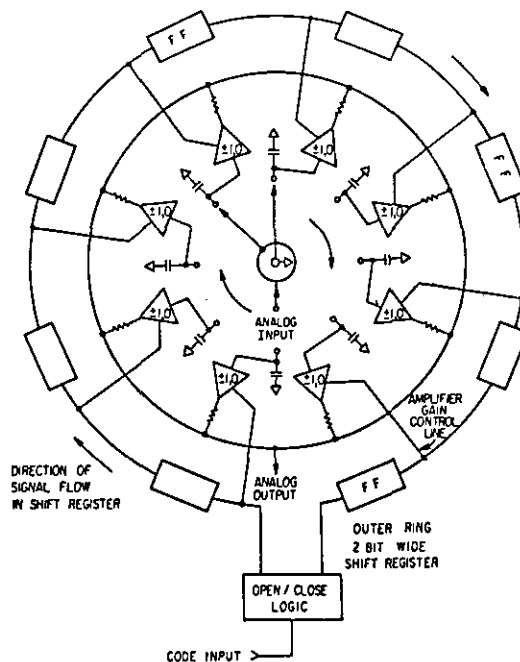

Figure 5. Block diagram of Analog Ring Decoder.

This sample block diagram depicts closely the way the actual machine operates. However, we can make a few simple additions which will allow the machine to run much faster with high accuracy. First let us make the number of capacitors and switch positions, M, somewhat larger than N, the length of the largest code we will use. Second, we arrange the phase of the switch rotor so that the code begins two positions after the switch pointer. This means that the present sample and the ones surrounding it will be connected to amplifiers with gain of 0. This allows the amplifier time to settle. Third we add a second rotor to the switch which runs one step ahead of the other. This rotor is connected to ground, and discharges each capacitor before it samples.

CONCLUSIONS AND RECOMMENDATIONS

MST radars should include hardware coherent integrators. This reduces the decoding efforts by many orders of magnitude. Once the decision to include a hardware coherent integrator has been made, a decoding operation can be included with little additional money effort. If coherent integration and decoding is performed with dedicated devices, the existing computing capacity can be used for statistical computations and parameter estimation.

Long sequences of codes, like the quasi-complementary sequences, can not be implemented with simple coherent integrators and decoders. Straight decoders are required, but devices simpler than the Arecibo decoder would be economical.

Bi-static CW radars should use continuous periodic pseudorandom codes. One bit correlators can perform the decoding operation economically.

REFERENCES

Barker, R. H. (1953), Group synchronizing of binary digital systems, in Communications Theory, W. Jackson (ed.), 273-287, Academic Press, New York

Barton, D. K. (ed.) (1975), Radar, vol. 3, Pulse Compression, Artech House.

Brookner, E. (ed.) (1977), Radar Technology, Artech House.

Cook, C. E. and M. Bernfeld (1967), Radar Signals: An Introduction to Theory and Applications, Academic Press, New York.

Golay, M. J. E. (1961), Complementary series, IEEE Trans. Info. Theory, IT-7, 82-87.

Gray, R. W. and D. T. Farley (1973), Theory of incoherent-scatter measurements using compressed pulses, Radio Sci., 8, 123-131.

Hagen, J., The Arecibo 1008 lag correlator. Use and Maintenance Manual - Arecibo, P.R.

Schmidt, G., R. Ruster and P. Czechowski (1979), Complementary code and digital filtering for detection of weak VHF radar signals from the meosphere, IEEE-GE.

Woodman, R. F. (1980), High-altitude-resolution stratospheric measurements with the Arecibo 2300 MHz radar, Radio Sci., 15, 423-430.

Woodman, R. F. (1980a), High-altitude resolution stratospheric measurements with the Arecibo 430-MHz radar, Radio Sci., 15, 417-422.

Woodman, R. F., R. P. Kugel and J. Rottger (1980), A coherent integrator-decoder preprocessor for the SOUSY-VHF radar, Radio Sci., 15, March.

13. THE ANALYSIS OF SPACED SENSOR RECORDS BY CORRELATION TECHNIQUES

B. H. Briggs

Department of Physics
University of Adelaide
Adelaide, Australia

INTRODUCTION

Several ground-based techniques used during the Middle Atmosphere Program involve the observation of moving patterns. Examples are:
1. The partial reflection drift method (PRD). (See G. J. Fraser, this Volume).
2. VHF radar using the spaced antenna wind method (SA). (See J. Rottger, this Volume).
3. Observation of airglow with spaced-field photometers.

Airglow intensities are recorded for three spaced regions in the sky, and motions derived from the recorded fluctuations.

The common feature of these observations (and others) is that it is difficult to record the entire moving pattern, and therefore the desired information about its motion and other properties must be derived rather indirectly from observations made at a few spaced points. A maximum of three spaced sensors is required.

There are many papers on this subject and we do not intend to give here a comprehensive review of the several different techniques of analysis which have been developed. The main object of the present paper is to present one particular method, based on correlation techniques, in a logical and simple way which it is hoped will be useful to those meeting the concepts for the first time. Some practical considerations which arise in the implementation of the method will also be discussed.

We begin by explaining some simple ideas about moving patterns, in order to explain why the correlation methods are needed.

SOME SIMPLE IDEAS ABOUT MOVING PATTERNS

It will often be useful to represent a pattern by <u>contours</u>, i.e. lines along which the measured quantity, whatever it may be, takes constant values. We will assume for simplicity that the measured quantity is real. (In the radio methods both phase and amplitude may be recorded and combined into a complex amplitude. This case is considered later in this paper. The pattern is assumed to be random in nature. and will therefore contain irregular "hills" and "valleys" if the measured quantity is regarded as being plotted perpendicular to the xy plane. The contours may be pictured as "contour lines" of these "hills" and "valleys". Figure 1 shows a typical "hill" moving with a speed V in a direction which makes an angle ϕ with the y-axis.

Suppose the measured quantity is recorded (as a function of time) by three sensors situated at O, X and Y. The sensors form a right angled triangle with separations ξ_0 in the x-direction and η_0 in the y-direction. To begin with we assume that the pattern moves past the sensors without change of form as it moves, and with a constant velocity. Then the sensor at O records a "section" through the "hill" along the dotted line O'O" which is drawn parallel to the direction of movement. Similarly, the sensors at X and Y record "sections" along the dotted lines X'X" and Y'Y".
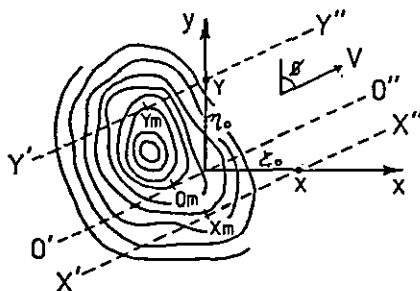
Figure 1.

The records obtained might look something like those of Figure 2. The maxima *correspond to the points marked* X_m, O_m, Y_m *in Figure 1, where the sections are* tangential *to the contours.*

A line joining $Y_m O_m X_m$ in Figure 1, is sometimes referred to as a line of maximum. If we had many more sensors we could define this line of maximum precisely. If the "hill" is large compared with the separation of the sensors O, X, Y the short segment of the "line of maximum" which passes over them may be taken as straight to a good approximation. It will also tend to be perpendicular to the direction of motion of the pattern. (This would be exactly true if the contours were circular, as the reader can easily verify; it will be true on the average if the contours are circular on the average. We reconsider this assumption later).

As a result of the above considerations we can replace the rather complicated diagram of Figure 1 by the much simpler one of Figure 3. Here a single "line of maximum" perpendicular to V is seen moving in the direction ϕ with a speed of V. It passes O first and then Y and X. We can calculate from Figure 3 the time differences t_x and t_y shown in Figure 2 as follows:-

$$t_x = \frac{\xi_0 \sin \phi}{V}, \quad t_y = \frac{\eta_0 \cos \phi}{V}.$$
(1)

Define $\quad v_x = \xi_0/t_x$ and $v_y = \eta_0/t_y$
(2)

We see that $\quad v_x = \dfrac{V}{\sin \phi}$, and $v_y = \dfrac{V}{\cos \phi}$.
(3)

It follows that v_x and v_y are not components of V.

To obtain V from v_x and v_y we can use the geometrical construction shown in Figure 4. A perpendicular from the origin to the line joining the v_x and v_y values gives the velocity vector, $\underline{V}$.

So far we have considered just one "hill" moving past the three sensors. We must now consider the succession of "hills" which move past. Since we assume the pattern to be of a random nature, each "hill" will have a different shape. The successive "lines of maxima" will be at different angles as sketched in Figure 5. It can be shown that they never depart greatly from the perpendicular to the line of motion, and that they are symmetrically distributed around this perpendicular direction provided the pattern has no systematic "elongation" in any direction. Therefore the average time displacements are unchanged, but the individual values $\bar{t}_x$, $\bar{t}_y$ will fluctuate. If average time displacements t_x and t_y are determined from the records as shown in Figure 6, the velocity can be found by calculating apparent "components"
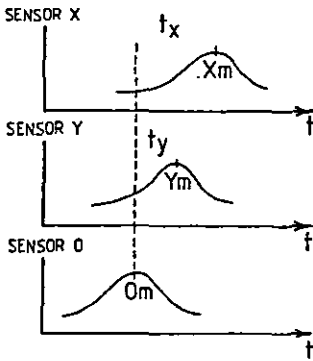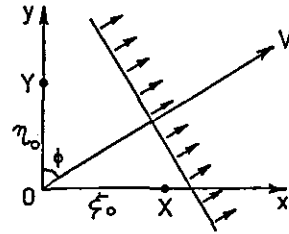
168

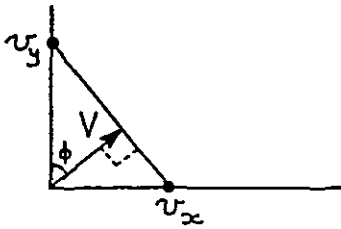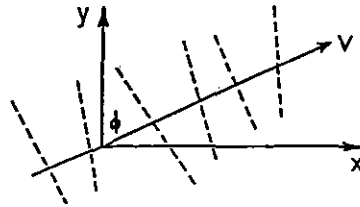Figure 2.

Figure 3.

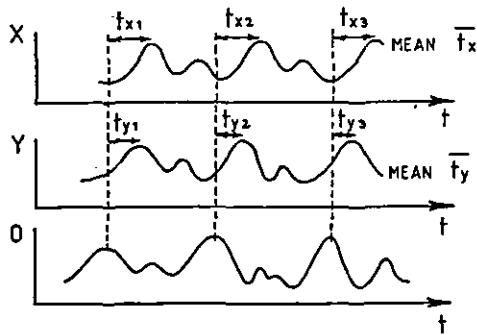

Figure 4.

Figure 5.

Figure 6.

$$\xi_0/\bar{t}_x \text{ and } \eta_0/\bar{t}_y \qquad\qquad (4)$$

and using the constuction of Figure 4.

In the early days of ionospheric drift analysis, $\bar{t}_x$ and $\bar{t}_y$ were often found by displacing the records relative to each other until they appeared most similar, rather than by measuring individual time displacements of particular features. Computing limitations prevented the determination of the mean time shifts by the cross-correlation technique to be described later.

At this point it is important to emphasize that the "lines of maxima" are not fixed features of the pattern. They depend upon its direction of motion and always tend to be perpendicular to this direction. "Lines of maxima" cannot be drawn for a stationary pattern.

With the assumptions made so far, the method described is a perfectly satisfactory method for determining the velocity of the pattern, and it may be asked why anything more elaborate is needed. The answer to this is that there are two factors, so far neglected, which could cause the calculated velocity to be seriously in error in some cases. These concern the fact that the pattern may be systematically elongated in a particular direction, and may change in form as it moves. We now discuss these effects.

If the contours are systematically elongated, as in Figure 7, the pattern will be called "anisometric". The lines of maxima, although still having some tendency to be perpendicular to the direction of motion, will be systematically displaced towards the direction of elongation. It is clear that the mean time displacements $\bar{t}_x$ and $\bar{t}_y$ will be changed and the true velocity cannot be found by the method of Figure 4. There may be a serious error in the calculated direction of motion if this effect is ignored.
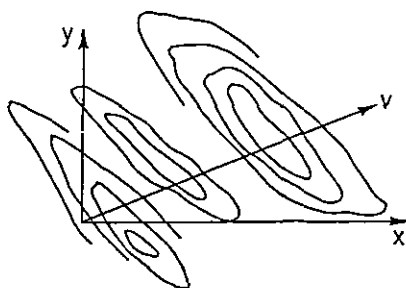

Figure 7.

It is likely that in many cases the pattern does not just move but also changes or 'evolves' in time as it moves. Errors in the calculated speed of movement will then occur. This can be seen by considering an extreme case: suppose there is no systematic movement at all, only random change of the pattern. The three records will still show similarity provided the sensors are separated by less than the scale of the pattern, but there will be no systematic time displacements. If individual time displacements t_x, t_y of particular features are measured, it will be found that $\bar{t}_x$ and $\bar{t}_y$ are both zero. Use of equations (4) will then lead to infinite apparent velocity components and an infinite velocity of movement of the pattern. More generally, if motion is combined with random change, the use of equations (4) will lead to values of the velocity which are too large. The direction will not be affected, provided the pattern is isometric (non-elongated).

The objectives of more elaborate methods of analysis are to eliminate these two sources of error. Two approaches have been used:

(1) Methods which still depend upon the determination of the individual time displacements t_x and t_y between corresponding features of the records, but which make use of the probability distributions and other properties of these quantities, to obtain better estimates of the velocity. As a simple example, one can calculate values of v_x and v_y from each separate pair of values of t_x and t_y using equation (2) and then average these to give $\bar{v}_x$ and $\bar{v}_y$. The velocity thus obtained will differ from that calculated from $\bar{t}_x$ and $\bar{t}_y$, and will be zero, rather than infinite, for a randomly changing but non-moving pattern. With a suitable theory, velocity estimates may be corrected, and other parameters of the pattern determined. A method based on determination of the individual time delays is called a "time-delay method", or "method of similar fades".

(2) The use of auto and cross-correlation functions rather than individual time delays, together with a theory which shows how these functions can be used to obtain the correct velocity and other parameters. This has become known as the method of "full correlation analysis".

In the present paper we will consider only method (2). Historically, method (1) was pursued first, because it was too laborious to obtain the correlation functions before the advent of computers. At the present time, method (2) seems more natural, since the calculation of correlation functions is well adapted to digital procedures, whereas the determination of time delays between maxima (for example) is less convenient and subject to ambiguities, since the computer has to be able to recognize which maxima correspond. However, we do not wish to argue this point here, and it should be mentioned that method (1) still has some advocates. Our main objective is to present the concepts of the correlation method in as simple a manner as possible. It is necessary to understand these concepts even if the time delay method (1) is preferred, since the mathematical formulation, and the definitions of the parameters are the same. (Readers who wish to study the time delay methods will find the main papers listed separately in the Bibliography).

CORRELATION ANALYSIS

In the above introduction, the idea of a randomly changing two-dimensional moving pattern was introduced in a rather descriptive and qualitative way. This idea needs to be expressed in a mathematical model before further progress can be made. It is important that the model does not introduce more parameters that can be determined experimentally. It will therefore be necessary to make some simplifying assumptions which may not always be true. The range of applicability of the model should be tested when it is applied to any particular data.

We will consider first the description of a stationary non-changing pattern. We will then introduce the random changes and finally the steady drift. In this way we arrive at the final model by a succession of simple stages.

(a) Description of a Stationary Non-Changing Random Pattern

Suppose first that the pattern is isometric (Figure 8), i.e. not systematically elongated in any preferred direction. Consider the expression

$$\rho(\xi,\eta) = \frac{<f(x,y) \cdot f(x+\xi, y+\eta)>}{<[f(x,y)]^2>} \tag{5}$$

This will be recognized as the standard statistical formula for the correlation co-efficient between the values of f(x,y) and of f(x+ξ, y+η), i.e. descriptively, between the pattern and the same pattern shifted by a distance ξ in the x-direction and η in the y-direction. If regarded as a function of ξ and η it is called the two-dimensional spatial correlation function of the pattern. It is unity when ξ = η = 0 and decreases as ξ and η increase, eventually tending to zero when the pattern is shifted so far that no correlation exists.

Two points about equation (5) need to be noted. First, the formula only gives the correlation coefficient for a function of zero mean value. The more general formula uses $(f(x,y) - <f(x,y)>)$ instead of f(x,y). As the resulting formula is rather cumbersome we will assume that $<f(x,y)> = 0$. With real data the full formula must be used for computations. Secondly, we must consider how the averaging indicated by the < > symbols is to be carried out. If we had two sensors S_1 and S_2 (Figure 10), these could be set at relative positions giving desired values of ξ and η and the pair then moved to all positions (x,y) in the pattern, keeping them in the same orientation and the same distance apart, (i.e. maintaining a constant vectorial separation $\underline{r}$ = (ξ,η)). For each position the product $f(x,y) \cdot f(x+ξ, y+η)>$ would be found and the average value of this product for all positions (x,y) would be used in equation (5). This procedure may be referred to as <u>spatial</u> averaging. The procedure would be repeated for all the desired values of ξ and η, thus determining the function ρ(ξ,η).

It will be realized that this procedure is not likely to be feasible in practice, and even if possible would be very laborius and time consuming. Fortunately, it is not necessary. When we introduce movements and random changes in time (as we will later) the sensors S_1 and S_2 can be placed at <u>fixed</u> positions with a vectorial separation $\underline{r}$, and the value of $<f(x,y) \cdot f(x+ , y+)>$ obtained as a <u>time</u> average. An "ergodic hypothesis" makes it reasonable to suppose that the results will be the same. Roughly speaking this just means that it does not matter what variables we average over, provided we obtain a random unbiassed set of values. However, even with time averaging the determination of ρ(ξ,η) would still be laborious if all values of (ξ,η) had to be sampled. If nothing is known about the form of ρ(ξ,η) this cannot be avoided. We can reduce the observational burden only if we are willing to make some assumptions. We now consider this point.

If the pattern is isometric the correlation should depend only on the magnitude of the displacement and not on the direction. That is, it should be a function of $|r| = (ξ^2+η^2)^{1/2}$. Contours of equal correlation (for which ρ(ξ,η) takes constant values) should therefore be circles in the (ξ,η) plane (Figure 9). We can picture this as a single circular "hill" of height unity at ξ = η = 0, which falls away to zero in the same way in all directions
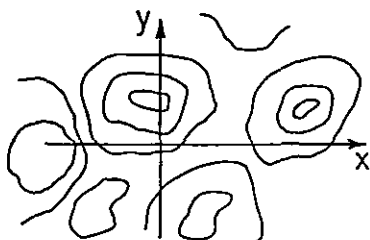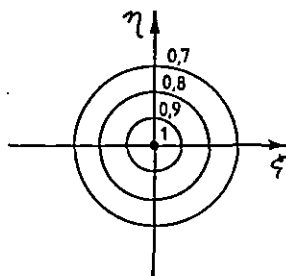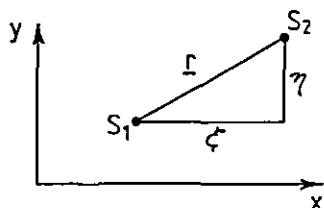

Figure 8.

Figure 9.

Figure 10.

from the origin. It is clearly enough if we know the "shape" of this hill in any one direction, e.g. the ξ-direction. Therefore sensors at different separations along the x-axis will be sufficient, either a pair whose separation is varied, or a fixed row.

If we have only two sensors at a fixed separation, ξ_0 along the x-axis, our information is even more limited. We can determine only one point on the spatial correlation function. This could nevertheless be useful if we wished to study how the pattern scale varied from occasion to occasion. If we adopted a particular functional form for the correlation function which contained only one variable parameter, then a single measurement at $\xi = \xi_0$ will fix the value of this parameter.

We must generalize these ideas to an anisometric pattern (Figures 11, 12). Again, if we had unlimited time and equipment, we could compute the contours of Figure 12 precisely by direct application of Equation (5). In practice, we must keep the number of parameters to be determined as small as possible. We therefore suppose that the anisometric pattern of Figure 11 has been produced by applying a simple "stretch" in a particular direction to the isometric pattern of Figure 8. Imagine the pattern of Figure 8 drawn on an elastic sheet which is then stretched. Only two parameters are then needed to specify the situation, the amount of stretch, and its orientation or direction relative to the (x,y) axes. The same transformation applied in the (ξ,η) plane converts the circles of equal correlation of Figure 9 into a family of similar ellipses (Figure 12). The axial ratio of these measurements the amount of stretch (anisometry or average elongation) and the orientation of their major axes gives the direction. The ellipse for a particular correlation value, usually $\rho(\xi,\eta) = 0.5$, is often referred to as the "characteristic ellipse". It sets the "scale" of the irregularities in the random pattern for each direction in the (x,y) plane.
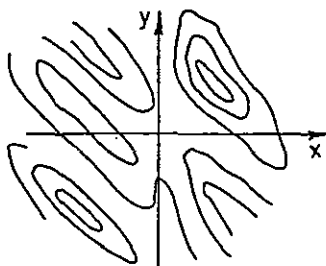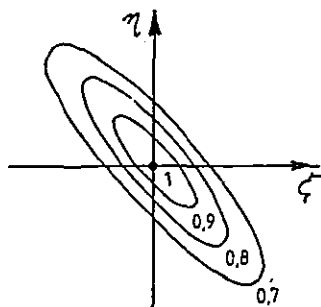

Figure 11.

Figure 12.

The assumption of elliptical contours puts certain mathematical constraints on the form the function $\rho(\xi,\eta)$ can take. If $\rho(\xi,\eta)$ = constant is to define an ellipse centered on the origin, the variables can only enter in the combination $(A\xi^2 + B\eta^2 + 2H\xi\eta)$. We therefore write

$$\rho(\xi,\eta) = \rho(A\xi^2 + B\eta^2 + 2H\xi\eta). \tag{6}$$

Note that the <u>form</u> of the function ρ is still arbitrary; the only assumption concerns the particular combination in which the variable ξ,η occur.

A STATIONARY BUT RANDOMLY CHANGING PATTERN

We now consider a pattern which has random changes in time but no tendency to move in any particular direction (rather like the surface of a boiling liquid). Mathematically, we now have a dependence on time, and we take the pattern as $f(x,y,t)$. A sensor at a fixed point, say the origin, would record time variations given by $f(0,0,t)$ and these could be described by a <u>temporal</u> auto-correlation function defined in the usual way. More generally, spaced sensors could be used to determine the function

$$\rho(\xi,\eta,\tau) = \frac{< f(x,y,t)\ f\ (x+\xi,\ y+\eta,\ t+\tau)>}{<[f(x,y,t)]^2>} \tag{7}$$

by making calculations in which time differences were introduced between the records obtained from the sensors with different separations.

Graphically we could represent the "evolving" pattern by the (x,y,t) diagram of Figure 13. Setting $f(x,y,t)$ = constant now gives <u>surfaces</u> in (x,y,t) space and the (x,y) plane intersects these surfaces to give the <u>contours</u> of Figures 8 or 11. As time advances, we must imagine the (x,y) plane to move upwards, as shown dotted, so that these contours change in a random way. Similarly, setting $\rho(\xi,\eta,\tau)$ = constant gives a family of surfaces in (ξ,η,τ) space, encircling the origin where $\rho = 1$ (Figure 14).

As before, to make progress we must make some assumptions about the form of these correlation surfaces. We first note that the features of Figure 13 cannot have any systematic tilt relative to the t axis; if they did, the spatial contours, which are cross-sections through them, would tend to move in the (x,y) plane in a preferred direction as time advances. This would be inconsistent with the assumption that the pattern has no systematic movement. This fact is reflected in Figure 14 by the correlation values always decreasing as we go away from the τ axis in any horizontal direction. Since there is no drift, introducing a ξ or η shift in addition to a τ shift can only result in a reduced correlation.
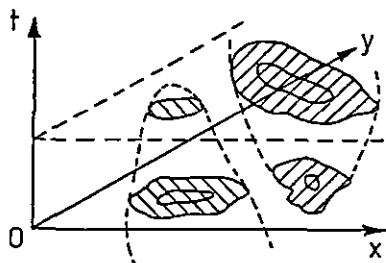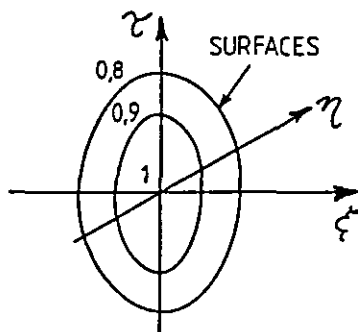

Figure 13.

Figure 14.

These considerations suggest that, by analogy with our previous assumptions about ellipses, the surfaces of Figure 14 might be represented by a family of concentric ellipsoids having one axis in the τ direction. These ellipsoids would intersect the (ξ,η) plane in a family of ellipses, as in Figure 12. That is, we might consider setting

$$\rho(\xi,\eta,\tau) = \rho(A\xi^2 + B\eta^2 + K\tau^2 + 2H\xi\eta). \tag{8}$$

No terms of the form $\xi\tau$ or $\eta\tau$ are permissible, since these would result in the family of ellipsoids being tilted with respect to the τ axis. From (8) surfaces of equal correlation are given by

$$A\xi^2 + B\eta^2 + K\tau^2 + 2H\xi\eta = \text{constant} \tag{9}$$

Putting $\tau = 0$, this gives Equation (6) for the spatial characteristic ellipses, as required.

It should now be stressed that the assumptions leading to Equation (8) are less plausible than those leading to Equation (6). With Equation (6) we were concerned with two spatial dimensions on an equal footing, and a first order representation of anisometry in the (x,y) or (ξ,η) planes is clearly a "stretch" in one preferred direction. However, in Equation (8) we are concerned with time as well as space, and it is not at all obvious that the way the pattern evolves in time is related in any simple way to the manner in which it varies spatially. The form (8) implies that the temporal autocorrelation function which would be obtained from a single sensor at one point in the pattern is given by

$$\rho(0,0,\tau) = \rho(K\tau^2), \tag{10}$$

and has therefore the <u>same functional form</u> as the spatial correlation function in any direction in space. For example, the spatial correlation function for sensors at separations ξ in the x direction is

$$\rho(\xi,0,0) = \rho(A\xi^2)$$

This assumption, that the spatial and temporal correlation functions have the same functional form, is never likely to be exactly true, and should be tested for each application of the theory to any experimental data.

(From a mathematical point of view, we could justify Equation (8) very close to the origin on the grounds that any function which has a maximum at the origin can be shown to be expressible in this form to the first order. However, this is not very helpful in practice because we will show later on that it is undesirable to work with very high values of correlation).

A MOVING AND RANDOMLY CHANGING PATTERN

We now suppose that the pattern described above moves with a speed V in a direction ϕ measured clockwise from the y-axis. For a pattern which is moving the surfaces $f(x,y,t) = \text{constant}$ will have a systematic slope as shown in Figure 15. As the (x,y) plane advances in the t direction the contours move systematically in the (x,y) plane. This is reflected in the (ξ,η,τ) diagram, in which the surfaces $\rho(\xi,\eta,\tau) = \text{constant}$ have a "tilt" relative to the τ-axis.

Mathematically, the introduction of the systematic motion is a relatively simple step. Consider an observer who "moves with the pattern". To this observer the pattern is only changing randomly, and Equation (8) must apply to any measurements this observer makes. We use primed co-ordinates (x',y',t)

for the moving observer, so that the pattern is f(x',y',t) and the associated correlation function is

$$\rho(\xi',\eta',\tau) = \rho(A\xi'^2 + B\eta'^2 + K\tau^2 + 2H\xi'\eta'). \tag{11}$$

A stationary observer records a pattern f(x,y,t), and his coordinates are related to those of the moving observer by the equations:

$$x = x' + Vt \sin\phi = x' + V_x t$$
$$y = y' + Vt \cos\phi = y' + V_y t$$

where $V_x = V\sin\phi$, and $V_y = V\cos\phi$, are the components of the systematic motion. It follows that

$$\xi = \xi' + V_x\tau \text{ and } \eta = \eta' + V_y\tau$$

Substituting in (11) we obtain

$$\rho(\xi,\eta,\tau) = \rho\{A(\xi-V_x\tau)^2 + B(\eta-V_y\tau)^2 + K\tau^2 + 2H(\xi-V_x\tau)(\eta-V_y\tau)\} \tag{12}$$

This is of the form

$$\rho(\xi,\eta,\tau) = \rho(A\xi^2 + B\eta^2 + C\tau^2 + 2F\xi\tau + 2G\eta\tau + 2H\xi\eta). \tag{13}$$

Equations (12) and (13) are the basic equations which will be used from now onwards. Note that surfaces $\rho(\xi,\eta,\tau) = $ constant still have the form of similar concentric ellipsoids but since terms in $\xi\tau$ and $\eta\tau$ are now present, these ellipsoids are no longer aligned along the τ-axis, but tilted as shown in Figure 16. The coefficients A,B,C,F,G,H, fully describe the situation, together with the functional form of ρ, which we emphasize once again is arbitrary and is to be determined from the data. We now show that this functional form, and the values of the coefficients, can be determined from experimental observations.

EXPERIMENTAL OBSERVATIONS

To determine a velocity with two independent components we expect to have to use a minimum of three sensors. We consider first the case in which these form a right-angled triangle and have separations ξ_0 in the x-direction, and η_0 in the y-direction, as in Figure 1.

Consider a cross-correlation function for the pair of sensors space along the x-axis. That is, experimentally, we make recordings at these two points, and then compute the correlation between them for different values of relative
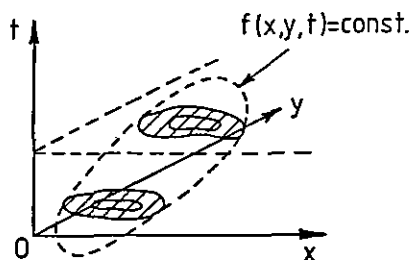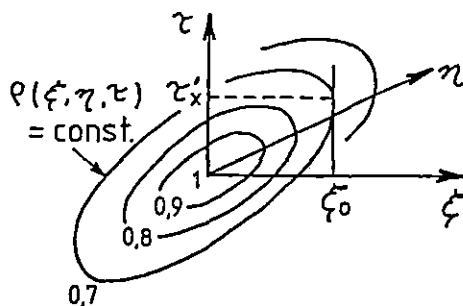

Figure 15.

Figure 16.

time shift τ. If the pattern is a moving one, we expect such a function to show a maximum for some value of time shift which we will call τ_x' (Figure 17). Mathematically, this function is obtained from the general equation (13) by the substitutions $\xi = \xi_0$, $\eta = 0$. Thus

$$\rho(\xi_0 0, \tau) = \rho\{A\xi_0^2 + C_T^2 + 2F\xi_0\tau\}.$$

To find the maximum, we set $\frac{\partial \rho}{\partial \tau} = 0$; this gives

$$\tau_x' = -\frac{F}{C}\xi_0. \tag{14}$$

(The reader should relate Figure 17 geometrically to Figure 16). Similarly, the time lag for maximum correlation for the sensors in the y-direction, separated by η_0, is given by
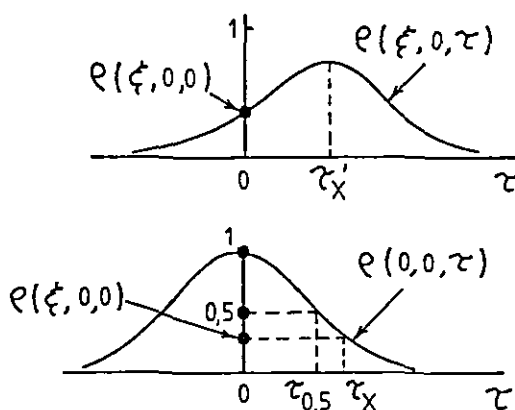
Figure 17.

$$\tau_y' = -\frac{G}{C}\eta_0. \tag{15}$$

Thus by measuring τ_x' and τ_y' we can find the values of F/C and G/C.

To proceed further we make use of the mean auto-correlation function. In theory the three auto-correlation functions for the three sensors should be the same. In practice statistical errors may introduce slight differences, but the mean of the three can be equated to the theoretical auto-correlation functions obtained from Equation (13), i.e. $\rho(0,0,\tau)$. This is

$$\rho(0,0,\tau) = \rho(C_T^2). \tag{16}$$

Suppose we measure $\rho(\xi_0,0,0)$ (the value of cross correlation at zero shift, i.e. just the simple correlation coefficient between the spaced records). We can read off the time shift τ_x at which the auto-correlation function has the same value (see Figure 17). So mathematically, τ_x satisfies the equation $\rho(\xi_0,0,0) = \rho(0,0,\tau_x)$. Using Equation (13) and (16) this gives $\rho(A\xi_0^2) = \rho(C\tau_x^2)$ and so

$$A/C = \tau_x^2/\xi_0^2. \tag{17}$$

Similarly, the other pair of sensors along the y-axis can be used to find the corresponding quantity τ_y and hence

$$B/C = \tau_y^2 / \eta_0^2. \tag{18}$$

We can similarly make use of the cross correlation between the pair situated at $(\xi_0, 0)$ and $(0, \eta_0)$. For zero time shift these have a correlation

$$\rho(\xi_0, \eta_0, 0) = \rho(A\xi_0^2 + B\eta_0^2 + 2H\xi_0\eta_0),$$

and the auto-correlation function takes the same value for a time shift τ_{xy} given by

$$\rho(A\xi_0^2 + B\eta_0^2 + 2H\xi_0\eta_0) = \rho(C\tau_{xy}^2).$$

This gives

$$\frac{H}{C} = \frac{\tau_{xy}^2}{2\xi_0\eta_0} - \frac{A \xi_0}{2C\eta_0} - \frac{B\eta_0}{2C\xi_0} \tag{19}$$

The value of H/C can be found from this equation since the ratios A/C and B/C have already been found. Thus the five ratios A/C, B/C, F/C, G/C, H/C have now been derived from the auto- and cross-correlation functions obtained from the three spaced sensors. These ratios fix the function (13) apart from the actual functional form of ρ. This is the same as the functional form of the observed average auto-correlation function $\rho(0,0,\tau)$ and is therefore known (Equation 16). To fix the "scale" it is necessary to measure one more parameter. This can be the value $\tau_{0.5}$ for which the auto-correlation function is equal to 0.5. i.e.

$$\rho(0, 0, \tau_{0.5}) = \rho(C\tau_{0.5}^2) = 0.5 \tag{20}$$

(see Figure 17).

We have illustrated the method of deriving the basic parameters of Equation (13) for the case of three sensors which form a right-angled triangle. This is, however, not the *optimum configuration in practice*. If three *sensors* are used, an equilateral triangle is preferable, as this more symmetrical arrangement is less likely to produce a bias in directional determinations. In general, more than three sensors may be used, in arbitrary positions.

Consider any number and arrangement of such sensors, and let the pair, i,j have a vectorial separation $\underline{r}_{ij}$, with an x-component separation ξ_{ij} and a y-component separation η_{ij}. Then the cross-correlation function for this pair. as a function of relative time shift τ_{ij} is given by

$$\rho(\xi_{ij}, \eta_{ij}, \tau_{ij}) = \rho(A\xi_{ij}^2 + B\eta_{ij}^2 + C\tau_{ij}^2 + 2F\xi_{ij}\tau_{ij} + 2 G\eta_{ij}\tau_{ij} + 2H\xi_{ij}\eta_{ij}) \tag{21}$$

The time shift for maximum correlation, τ_{ij}', can be found by setting $\partial\rho/\partial\tau_{ij} = 0$ as before, which gives

$$\tau_{ij}' = -\frac{F}{C} \xi_{ij} - \frac{G}{C}\eta_{ij} \tag{22}$$

Thus, the values of τ_{ij} obtained from each sensor pair can be used to obtain values of F/C and G/C, by solving a set of simultaneous linear equations. In general these coefficients will be "over-determined", and optimum values, as well as estimates of errors, can be obtained using a least squares formulation.

The cross-correlation between the i,j pair at zero time shift is $\rho(\xi_{ij}, \eta_{ij}, 0)$. We can find the time shift τ_{ij} at which the mean auto-correlation function (averaged over all sensors) has this same value. Then

$$\rho(0,0,\tau_{ij}) = \rho(\xi_{ij},\eta_{ij},0)$$

From Equation (13) this gives

$$\rho(C\tau_{ij}^2) = \rho(A\xi_{ij}^2 + B\eta_{ij}^2 + 2H\xi_{ij}\eta_{ij})$$

and so

$$\tau_{ij}^2 = \frac{A}{C}\xi_{ij}^2 + \frac{B}{C}\eta_{ij}^2 + \frac{2H}{C}\xi_{ij}\eta_{ij} \tag{23}$$

Again, the coefficients A/C, B/C, H/C will in general be over determined, and optimum values and errors can be found by a least square technique, using all the available values of τ_{ij}. This illustrates how the parameters of Equation (13) can be found for an arbitrary configuration of three or more sensors.

We have described above how the basic function of Equation (13) may be derived from the observations. It is simple and straightforward, but not necessarily the best method. It may be criticized on the grounds that only a few points on the experimental correlation functions are used, and the rest of the information is discarded. A preferable method (but see text under "Some Practical Considerations") might be to use a fitting procedure, and to vary the parameters A to G in order to obtain a best fit (on a least squares error basis) to the complete set of experimental cross-correlation functions. The form of ρ would be taken to be the same as the mean auto-correlation function.

We will assume that Equation (13) has been derived from the experimental data by one method or another. The following discussion is quite unaffected by the particular method by which this was done.

CALCULATION OF USEFUL QUANTITIES

The basic Equation (13) fully describes the situation, but not in terms of directly useful quantities such as the velocity of the pattern. We must now see how the physically significant quantities can be obtained.

CALCULATION OF THE VELOCITY

The components of the mean velocity are V_x and V_y. These can be found by equating the coefficients of $\xi\tau$ and $\eta\tau$ in Equations (12) and (13). This gives

$$AV_x + HV_y = -F \tag{24}$$
$$BV_y + HV_x = -G$$

If these equations are divided through by C, we have a pair of simultaneous equations for V_x and V_y in which the coefficients A/C, H/C, F/C, B/C, G/C are known. Hence V_x and V_y can be found. Note that V_x and V_y are true components of the velocity V. The magnitude and direction of V are therefore given by

$$|V|^2 = V_x^2 + V_y^2 \tag{25}$$

and

$$\tan \phi = V_x/V_y. \tag{26}$$

(However, it is often convenient to leave the results in component form).

The vector velocity $\underline{V}$ determined as above has become known as the "true" drift velocity to distinguish it from the "apparent" drift velocity determined from the mean time shifts as described in the introduction. That is, $\underline{V}$ has been corrected for anisometry of the moving pattern, and for the effect of the random changes of the pattern. It is important to put the word "true" in inverted commas however, since the corrections involve assumptions which may not hold in practice, and which should be tested, if possible, by comparisons with some more direct method of determining the velocity.

DESCRIPTION OF THE SPATIAL PROPERTIES OF THE PATTERN

The scale and degree of anisometry of the pattern are often of direct interest in themselves.

The spatial correlation function is

$$\rho(\xi, \eta, 0) = \rho(A\xi^2 + B\eta^2 + 2H\xi\eta). \tag{27}$$

Contours of equal correlation are given by the family of ellipses

$$A\xi^2 + B\eta^2 + 2H\xi\eta = \text{const.} \tag{28}$$

The orientation of the major axes, and the axial ratio (major/minor axis length) can be determined easily from Equation (28) using methods given in textbooks of co-ordinate geometry.

The particular ellipse for which $\rho = 0.5$ may be defined as the "characteristic ellipse". As we have already explained, it describes the "scale" of the irregularities in the pattern in different directions. This ellipse is defined by

$$\rho(A\xi^2 + B\eta^2 + 2H\xi\eta) = 0.5 \tag{29}$$

Comparing this with Equation (20) (which was used to define $\tau_{0.5}$) we obtain

$$A\xi^2 + B\eta^2 + 2H\xi\eta = C\tau_{0.5}^2. \tag{30}$$

as the equation of the characteristic ellipse. After dividing through by C, it is clear that all the coefficients are known.

The spatial properties of the pattern are usually specified by giving the values of the minor axis of the characteristic ellipse, the axial ratio, and the orientation of the major axis, measured clockwise from north.

DESCRIPTION OF THE RANDOM CHANGES

Although the physical significance of the random changes may be difficult to determine in any specific application (see the next section), a statistical measure of the importance of such changes is needed if our description of the moving pattern itself is to be complete.

For an observer who moves with the drift velocity of the pattern, the only time changes observed will be those due to the random changes or "evolution" of the pattern as it moves. These changes may be described by the temporal auto-correlation function which such an observed would obtain by analyzing the fluctuations recorded by a single sensor. From Equation (11) this auto-correlation function is

$$\rho(0, 0, \tau) = \rho(K\tau^2). \tag{31}$$

It is sometimes useful to describe the random changes in terms of a "mean lifetime" or "time-scale". Conventionally this is often taken to be the time lag $T_{0.5}$ at which the auto-correlation function falls to 0.5, i.e.

$$\rho(0,0,T_{0.5}) = \rho(KT_{0.5}^2) = 0.5 \tag{32}$$

We have already defined a quantity $\tau_{0.5}$ which is the time lag for which the directly observed auto-correlation function (for a fixed observed) falls to 0.5. Comparing Equations (20) and (32), we obtain

$$T_{0.5} = \tau_{0.5} \frac{C}{K}. \tag{33}$$

To find C/K we equate the coefficients of τ^2 in Equations (12) and (13). This gives

$$C = AV_x^2 + BV_y^2 + K + 2HV_xV_y \tag{34}$$

Since V_x and V_y, and the ratios A/C, B/C, H/C are known, this equation can be used to find C/K. Then $T_{0.5}$ can be found from (33), as $\tau_{0.5}$ is directly observable.

The description of the random changes in terms of a lifetime or time-scale is the most natural one, but historically, a different description was used. A velocity V_c was defined as follows. Consider again the observer moving with the pattern. Such an observer can measure the spatial correlation function directly. Suppose for the moment that the pattern is isometric. It can then be described by a one-dimensional spatial correlation function obtained from sensors arranged along a line. The same observer can also calculate the temporal auto-correlation function from records at a single sensor, as just described, and this is entirely due to the random changes. With the assumptions we have made, these correlation functions will have the same shape (functional form) and can therefore be related by a constant scale factor. Since on function is spatial and the other temporal, the scale factor has dimensions of velocity. This velocity is V_c; formally it may be defined as the ratio of the space separation to the time lag for equal values of correlation. It may be regarded as the ratio of the "irregularity size" to the "irregularity lifetime", for an observer moving with the pattern.

If the pattern is anisometric, the pattern scale depends on direction and V_c, as defined above, also depends on direction. The value of V_c appropriate to the direction of motion has special significance, as we will see. This quantity, denoted by $(V_c)_v$, is defined as the irregularity size in the direction of motion, divided by the irregularity lifetime. Thus if $d_{0.5}$ is the separation, in the direction of motion, for which the correlation between two sensors falls to 0.5 we define

$$(V_c)_v = \frac{d_{0.5}}{T_{0.5}} \tag{35}$$

To obtain the value of $d_{0.5}$ we note that the spatial correlation function is given by Equation (27). If d is a distance variable measured in the direction of the velocity V we have

$$\frac{\xi}{V_x} = \frac{\eta}{V_y} = \frac{d}{V}. \tag{36}$$

Substituting in (27) we obtain

$$\rho(d) = \rho\{(AV_x^2 + BV_y^2 + 2H\ V_xV_y)\ d^2/V^2\}. \tag{37}$$

Thus $d_{0.5}$ is obtained by setting

$$\rho\{(AV_x^2 + Bv_y^2 + 2H\ V_xV_y)\ d_{0.5}^2/V^2\} = 0.5 \tag{38}$$

Comparing this with Equation (32) which was used to define $T_{0.5}$ we obtain

$$(AV_x + BV_y + 2H\ V_xV_y)\ d_{0.5}^2/V^2 = KT_{0.5}^2,$$

and using $V^2 = V_x^2 + V_y^2$ we find

$$(V_c)_v^2 = \frac{d_{0.5}^2}{T_{0.5}^2} = \frac{(V_x^2 + V_y^2)K}{(AV_x^2 + BV_y^2 + 2H\ V_xV_y)}\ . \tag{39}$$

The value of $(V_c)_v$ can be found from this equation since V_x, V_y and the ratios A/K, B/K, H/K are known.

In order to appreciate fully the contributions of drift and random changes to the fluctuations recorded by a single fixed sensor it is instructive to consider further the form of the temporal auto-correlation function which is obtained from such a sensor. This is obtained from Equations (12) and (13) as $\rho(0,0,\tau)$ i.e.,

$$\rho(0,0,\tau) = \rho(C\tau^2) = \rho\{(AV_x^2 + BV_y^2 + 2H\ V_xV_y + K)\ \tau^2\} \tag{40}$$

The first three terms represent the time changes produced by the translation of the pattern past the fixed sensor with velocity V. This can be seen by making the substitution $d = V\tau$ in Equation (37), which gives

$$\rho\{(AV_x^2 + BV_y^2 + 2H\ V_xV_y)\ \tau^2\}\ .$$

If there were no random changes this would be the only mechanism for producing time changes at a fixed point, and the observed temporal auto-correlation function would necessarily be of the same functional form as the spatial correlation function.

On the other hand, if there is no mean movement ($V = 0$) and only random changes, we have already seen that the temporal auto-correlation function would be $\rho(K\tau^2)$ (as for an observer moving with the pattern). Equation (40) shows how the two effects are "combined" or "added" on the assumption that the random changes are described by an auto-correlation function which has the same shape as the spatial correlation function. The ratio of the two contributions to fading at a fixed point depends on the quantity

$$\frac{K}{AV_x^2 + BV_y^2 + 2H\ V_xV_y} = \frac{(V_c)_v^2}{V^2} \tag{41}$$

(using Equation 39). Thus the dimensionless parameter $(V_c)_v/V$ may be regarded as a measure of the <u>relative importance of random changes, compared with movement,</u> in producing the time variations at a fixed point. If this parameter is zero, the time changes are entirely due to the translation of fixed irregularities past the observer. If the parameter is infinite, the time changes are entirely due to random changes, with no contribution from motion. This interpretation gives additional significance to the random velocity parameter $(V_c)_v$. Also Equation (40) shows that our definition of the motional velocity V is effectively the following: the velocity V is the velocity of an observer who so adjusts his motion as to obtain the most slowly decreasing temporal auto-correlation function, i.e. the slowest possible rate of fading.

PHYSICAL SIGNIFICANCE OF V_c.

In connection with the radio experiments there has been much discussion as to whether V_c has any direct physical interpretation. It is tempting to try to relate it to the turbulent velocity in the scattering region. Some words of caution are necessary here.

Firstly, although V_c has dimensions of velocity, it need not represent any physical velocity; it is simply the ratio of the size $d_{0.5}$ to the life-time $T_{0.5}$ of the irregularities in the pattern. We could imagine a random pattern in which irregularities simply appeared and then decayed without moving around during their life. For such a pattern V_c would have a definite value, but could not in any sense be regarded as a velocity of motion. The opposite extreme would be a pattern in which the features moved around randomly but did not decay. Since $T_{0.5}$ is a measure of the time-scale at a fixed point in the pattern it would have a definite value, and it is easy to see that in this case V_c could be regarded as a measure of the random velocities. It is possible to imagine cases between these two extremes.

Even in cases where V_c does measure the velocity of random motions in the pattern, it does not follow that it is any direct measure of the motions of the scattering centers observed by the radar, because the connection between the pattern on the ground and the scattering irregularities is very indirect. Consideration of this connection suggests that the life-time or time-scale $T_{0.5}$ is a more physically significant quantity than is V_c.

THE APPARENT VELOCITY

The time delays τ_x' and τ_y' for maximum cross-correlation between the "x-pair" of sensors and the "y-pair" of sensors are closely analogous to the mean time shifts $\bar{\tau}_x$ and $\bar{\tau}_y$ of Equation (4). They can be used in the same way to calculate "apparent velocity components" which in turn lead to an "apparent velocity", V_a, using the construction of Figure 4. This velocity will in general be too large and its direction ϕ_a will be incorrect due to the neglect of random changes and anisometry.

The reader may have doubts about the value of calculating the "apparent" velocity V_a, since the full correlation analysis claims to be able to calculate the "true" velocity V. This point would be valid if we had complete confidence in the assumptions of the correlation analysis. In practice we may not have this confidence in the "corrections" for random changes and anisometry, especially when these corrections are large. It is therefore desirable to find V_a so that V and V_a may be compared. If V and V_a are very different for any particular record, it may be advisable to reject the data for that sample, unless the assumptions have been very thoroughly tested, or the values of V checked by some independent method, over a wide range of the ratio V_a/V.

SOME PRACTICAL CONSIDERATIONS

(a) Introduction

The implementation of the full correlation analysis can be carried out with various degrees of elaboration. The most sophisticated approach would be a complete least square fit of all the correlation functions for a large assembly of sensors to a model which took the shape of the mean auto-correlation function as the basic functional form. Although such a procedure has been suggested it does not appear to have been used for the production of routine data. For radio observations of upper atmosphere winds the emphasis has been,

rather, to use a simple a form of the analysis as possible, so that it can be implemented in real time by an on-line computer. This argues for the use of three sensors only, and an application of an approach similar to that outlined in this paper. The main computation time then comes from the evaluation of the auto- and cross-correlation functions; the rest is straightforward algebra.

The following sections deal with various practical considerations which have arisen in applying the method to wind determinations using radars at MF and VHF, using three spaced antennas as the sensors (SA method).

(b) Antenna Spacing and Record Length

The optimum configuration for the three antennas is an equilateral triangle with a spacing such that the cross-correlation between the antenna pairs at zero lag is of the order of 0.5. Comparisons with winds obtained by other methods have shown that if the correlation is appreciably higher than this value (sensors too close) the "true" velocity tends to be too small. On the other hand, if the correlation is too small (sensors too far apart) the analysis frequently breaks down.

At the frequencies under consideration, the fading period is of the order of one second. Therefore, records as short as one minute will suffice to produce statistically reliable correlation functions, giving excellent time resolution.

(c) Interpolation Procedures and Error Estimates

The correlation functions will be computed at discrete points, and to obtain the values of τ_{ij}' at which the cross-correlation functions are a maximum interpolation will be necessary. An interpolation polynomial may be fitted to five data points near the central portion of the peak. A polynomial of second degree is usually adequate. Similar interpolation procedures can be used to find the values τ_{ij}.

The errors in the τ_{ij} values are likely to be much less than in the τ_{ij}' values, since the former are determined from steeply sloping curves, while the latter depend on the determination of maxima. We therefore neglect errors in τ_{ij} and consider the probable errors of the τ_{ij}'.

BUCKLEY (private communication) has shown that the standard deviation σ of τ_{ij}' is given approximately by

$$\sigma = \tau_{0.5} \sqrt{\frac{1 - \rho_m^2}{\rho_m}} \sqrt{\frac{\tau_{0.5}}{T}} \tag{42}$$

Here $\tau_{0.5}$ is the half-width of the correlation functions, ρ_m is the maximum value of ρ_{ij} (from which τ_{ij}' is determined) and T is the record length. This formula has an intuitively reasonable form; we would expect the error to be proportional to the width of the correlation function, and inversely proportional to the square root of the number of independent data points N (where N ~ $T/\tau_{0.5}$). The factor involving ρ_m ensures that the error is zero if $\rho_m = 1$ and infinite if $\rho_m = 0$.

Having found the errors in the τ_{ij}', this leads to errors in the parameters A, B etc., and then to errors in the final velocity estimates.

If external noise is significant, further errors may arise because of the uncertainties in the method of "spike" removal described in the following section.

(d) Noise Removal

If the data contains high frequency noise, the auto-correlation functions will not extrapolate smoothly to unity at zero time shift, but will contain a "spike" at the origin. The cross-correlation functions may or may not have spikes at zero lag, depending on the type of noise. If the noise is correlated at the antenna pairs (e.g. external man-made noise, or atmospherics) spikes will be present. If the noise is not correlated (e.g. receiver noise or galactic noise) there will be no spikes.

If spikes are present, they may be removed as follows. Firstly, remove any spikes on the cross-correlation functions by straightforward interpolation. Secondly, determine the interpolated maximum value of each auto-correlation function at zero lag by fitting a polynomial function to points on each side of the maximum. This interpolated value, S, at zero lag, is a measure of the mean square signal. The value $N = 1-S$ is the mean square noise component. Thus the signal to noise ratio S/N can be found. The effect of the noise is then removed by re-normalizing the auto-correlation function to unity at the origin, which involves (for antenna i) dividing by S_i. The mean auto-correlation function can then be found as the average of the corrected auto-correlation functions for all the antennas. The cross-correlation functions must also be re-normalized. This involves dividing the covariance product for the pair i,j by the factor $(S_i S_j)^{1/2}$, where S_i and S_j are determined from the auto-correlation functions as just described.

(e) Use of Phase and Amplitude

In the radio experiments, the pattern which moves past the antennas is a pattern of amplitude and phase. In early work, only the amplitude was recorded, using simple peak detectors in the receivers. In this case, the formulae for the correction functions given earlier are directly applicable. More recently, both phase and amplitude have been recorded, in order to make fuller use of the available information. This involves the use of phase-sensitive detectors, recording in-phase and quadrature components of the received signals. These can be combined into a single complex quantity in the usual way after subtracting the mean value from each. When the variable is complex, the formulae for auto- and cross-correlation functions are modified by using the complex conjugate of one of the multiplied terms in the numerator. The resulting functions are in general complex. The modulus of the computed functions can be used to apply the full correlation analysis, which then proceeds as before.

The use of phase as well as amplitude has the advantage that coherent integration of several successive incoming pulses may be used. (The pulse rate is usually considerably higher than the required data sampling rate). If m successive pulses are averaged at the outputs of the in-phase and quadrature detectors, the signal to noise power ratio is increased by a factor m. This compares with a factor $\sqrt{m}$ which would apply if only amplitude detectors were used, and m successive pulses were averaged.

(f) Rejection Criteria

It is usually found that some records and/or results must be rejected, either because of inadequacy of the data itself (e.g. weak signal), or because the data does not appear to satisfy the assumptions of the analysis. The criteria used will of course depend upon the nature of the data. The following criteria have been found satisfactory for data relating to mesospheric winds observed at Adelaide on a frequency of 2 MHz, and may be a useful guide for other observations of a similar type. Some of them have been mentioned earlier, but it will be convenient to collect them together at this point. Data and/or results are rejected if any of the following conditions occur:

(1) The mean signal is either very weak or so strong that the receivers are saturated most of the time.

(2) The fading is very shallow (signal almost constant). The standard deviation should be at least 2% of the mean signal.

(3) The signal to noise ratio (obtained as described in Section d) is less than − 6 dB.

(4) The mean auto-correlation function has not fallen to at least 0.5 for the maximum number of lags which have been computed. (This condition will occur if the fading is unusually slow).

(5) The cross-correlation functions have no maxima within the number of lags which have been computed.

(6) The cross-correlation functions are oscillatory, so that it is impossible to identify the correct maxima.

(7) The sum of the time displacements τ_{ij} does not sum to zero, as it ideally should, for the data taken in pairs around the triangle of antennas. This condition may be relaxed somewhat to give a more practical criterion: If

$$\frac{|\Sigma \ \tau_{ij}|}{\Sigma \ |\tau_{ij}|} > 0.2$$

the results are rejected. This quantity has been called the "normalized time discrepancy". ·

(8) The computed value of V_c^2 is negative, so that V_c is imaginary. This criterion should not be applied too rigidly, because if the compute V_c^2 is only slightly negative, this probably means that the true value is zero, and the small negative value has arisen from statistical fluctuations. Such data may therefore be particularly good, indicating pure drift with negligible random changes.

(9) The computed coefficients indicate hyperbolic rather than elliptical contours, i.e. reject, if $H^2 < AB$

(10) The polynomial interpolation procedures break down at any stage.

(11) It is advisable to reject any results for which the corrections of the full correlation analysis are very large, i.e. those for which the apparent and true velocities differ greatly in magnitude and or direction. Suitable criteria are: Reject if

$$V_a \geq 3V \quad \quad or$$

$$V_a \leq 0.5V \quad \quad or$$

$$(\phi_a - \phi) \geq 40^0$$

REFERENCES

Papers on the basic techniques of correlation analysis

Baker, P. W. (1981), Estimation of the parameters of a turbulent drifting medium by a least squares analysis of spaced sensor measurements, Radio Sci., 16, (2), 213-217.
Briggs, B. H., G. J. Phillips and D. H. Shinn (1950), The analysis of observations on spaced receivers of the fading of radio signals, Proc. Phys.

Soc. London, B63(2), 106-121.

Brown, G. M. and J. W. Chapman (1972), Full correlation ionospheric drift analysis for a general observing triangle. Ann. Geophys., 28(2), 349-356.

Chandra, H. and B. H. Briggs (1978), The effect of filtering on ionospheric drift parameters determined by full correlation analysis, J. Atmos. Terr. Phys., 40(59), 541-548.

Fedor, L. S. (1967), A statistical approach to the determination of three-dimensional ionospheric drifts. J. Geophys. Res., 72(21), 5401-5415.

Gusev, V. D. and S. F. Mirkotan (1960), 3D full correlation analysis (in Russian) Issl. Neodnorod. ionos., 4, 7-19, Izd. Akad. Nauk. Moskva.

Meek, C. E. (1980), An efficient method for analyzing ionospheric drifts data, J. Atmos. Terr. Phys., 42(9/10), 835-839.

Meek, C. E., A. H. Manson and J. B. Gregory (1979), Internal consistency analysis for partial and total reflection drifts data, J. Atmos. Terr. Phys., 41(3), 251-258.

Phillips, G. J. and M. Spencer (1955), The effects of anisometric amplitude patterns on the measurement of ionospheric drifts, Proc. Phys. Soc. London, B68(8), 481-492.

Yerg, D. G. (1955), Notes on correlation methods for evaluating ionospheric winds from radio fading records, J. Geophys. Res., 60(2), 173-185.

Papers on time delay methods (method of "similar fades")

Bamgboye, D. K., A. J. Lyon and R. M. Morriss (1974), Anisometry parameters and corrections using similar fades technique, J. Geophys. Res., 79(22), 3231-3233.

Banerji, R. B. (1958), Method of measuring winds by fading at spaced receivers, J. Atmos. Terr. Phys., 12(4), 248-257.

Briggs, B. H. and M. Spencer (1954), The variability of time shifts in measurements of ionospheric movements, Rep. Phys. Soc. Conf. on Physics of Ionosphere, 123-125.

Harnischmacher, E. (1968), Results of an experiment concerning the fading pattern. In "Winds and turbulence in stratosphere, mesosphere and ionosphere", Ed. K. Rawer, North Holland Pub. Co., 277-297.

Kurschner, D. and R. Schminder (1980), Forschritte bei der Algorithmierung and Standardisierung der automatischen Auswertung von Ionospharendrift-messungen im Langwellenbereich und ihre Bedeutung fur den Aufbau von Messnetzen zur synoptischen Analyse hochatmospharischer Windfelder, Geologie und Geophysik. Geophys Veroff, Karl Marx Univ. Leipzig, Akad. Verlag Berlin, Serie 3, Band 2, Heft 2, 219-227.

Mitra, S. N. (1949), A radio method of measuring winds in the ionosphere, Proc. Instn. Elect. Engrs., 96, Part III, No. 43, 441-446.

Putter, P. (1954), Messung des Ionospharenwindes aus der Wanderungsgesch-windingkeit eines Zustandes (z. B. Echo-Feldestarke) langes der Erdoberflache, Rep. Phys. Soc. Conf. on Physics of Ionosphere, 191-201.

Schminder, R. and D. Kurschner (1983), This volume.

Sprenger, K. and R. Schminder (1969), On some relationships between correlation analysis and similar-fade analysis results on drift measurements in the lower ionosphere, J. Atmos. Terr. Phys., 31(8), 1085-1098.

14. THE MST RADAR TECHNIQUE

J. Röttger*

EISCAT Scientific Association
Kiruna, Sweden

ABSTRACT

This tutorial review gives a general overview on the coherent radar technique with special emphasis to mesosphere-stratosphere-troposphere (MST) radars operating in the VHF band. Some basic introduction to Doppler radar measurements and the radar equation is followed by an outline of the characteristics of atmospheric turbulence, viewed from the scattering and reflection processes of radar signals. MST radar signal acquisition and preprocessing, namely coherent detection, digital sampling, pre-integration and coding, is briefly discussed. The data analysis is represented in terms of the correlation and spectrum analysis, yielding the essential parameters: power, signal-to-noise ratio, average and fluctuating velocity and persistency. The techniques to measure wind velocities, viz. the different modes of the Doppler method as well as the spaced antenna method are surveyed and the feasibilities of the MST radar interferometer technique are elucidated. A general view on the criteria to design phased array antennas, used in most MST radar systems, is given. An outline of the hardware of a typical MST radar system is presented, which consists of the transmitter-receiver part, the radar controller and the data acquisition unit. In conclusion, some typical results are depicted, presently planned and operational radars are surveyed and a brief outlook for future requirements for MST radar systems is attempted.

1. INTRODUCTION

The origin of MST radars dates back to the early days of ionospheric backscatter observations which were carried out with a 41 MHz radar in Illinois/USA by BOWLES (1958). Strong echoes, observed at 75-90 km height, were interpreted as due to "ionospheric scattering of the turbulence variety" in the mesosphere. Subsequently, FLOCK and BALSLEY (1967) reported about VHF radar observations at Jicamarca/Peru, which confirmed echoes from 75 km height and were also interpreted to be most likely caused by turbulence scattering in the presence of a gradient of electron density. Several years later Ronald F. WOODMAN with Alberto GUILLEN (1974) substantially improved the technique in order to measure velocities and they also reported the detection of strato-spheric returns. They recognized the great potential of this technique for remote sounding of the middle atmosphere. Their studies triggered the evolution of a new generation of radars for atmospheric research, and VHF radars were started thereafter to be developed for the only purpose of lower and middle atmosphere observations. Although earlier expectations that the entire region of the mesosphere, stratosphere and troposphere, which gave the name "MST" radar, could be monitored more or less continuously, have emerged to be an overestimate, their very unique applications to investigate the structure and dynamics of the middle atmosphere is widely accepted (GAGE and BALSLEY, 1978; WALKER, 1979; BALSLEY and GAGE, 1980; CRANE, 1980; HARPER and GORDON, 1980; JAMES, 1980; ROTTGER, 1980a; KLOSTERMEYER, 1981; RASTOGI, 1981; WOODMAN, 1981; LARSEN and ROTTGER, 1982; LARSEN, 1983a). It has turned out that the MST radar technique is also very suitable for operational applications in meteorology (LHERMITTE, 1979; WILSON et al., 1980; ROTTGER, 1981a; BALSLEY and

---

* Presently at Arecibo Observatory, Arecibo, Puerto Rico; on leave from Max-Planck-Institut für Aeronomie, W. Germany.

GAGE, 1982; HOGG et al., 1983; LARSEN 1983b). It is, thus, expected that further useful information on atmospheric physics will be obtained from continued MST radar observations, and these will also gain substantial contributions for the Middle Atmosphere Program.

MST radars make use of scattering and reflection from variations of humidity, temperature and electron density, induced by turbulence in the lower and middle atmosphere. Essentially, MST radars can observe: the 3-dimensional wind vector, atmospheric reflectivity and stability, and morphology of turbulence and waves. The continuous measurements with MST radars offer very good quality and quantity middle atmosphere observations of wind velocities (GAGE and VANZANDT, 1981). MST radars operate at frequencies around 50 MHz, and therefore are also called VHF radars (VHF = very high frequency band between 30 MHz and 300 MHz). Higher frequency radars mostly cover only the troposphere and stratosphere. Typical peak powers of VHF radars are between 1 kW and 1 MW. Range resolutions down to about 100 m and time resolutions down to some ten second are possible. The antenna arrays with typical dimensions of 1000 m^2 to some 10,000 m^2 point close to the zenith direction. Coherent detection, digital control and data acquisition are mandatory. We notice that the instrumental technique as well as the data acquisition and analysis are now basically developed, elaborated and fairly mature, although further refinement is necessary.

This tutorial review aims to give a general summary and overview on the state-of-the-art achieved during the first decade of the MST radar technique. It is intended to yield also a brief introduction to the atmospheric radar technique with special emphasis to MST radar applications. It is not intended to represent a review for experts working in this field, but rather to give an overview to outsiders as well as to allow a basic introduction for those who are planning to join this research field. For more specific technical descriptions of the atmospheric radar technique, the reader is referred to other books and articles, such as those by SKOLNIK (1970), HARDY (1972), WILSON and MILLER (1972), BATTAN (1973), GOSSARD and STRAUCH (1983), DOVIAK and ZRNIC (1984). Very detailed descriptions of MST radar techniques can be found in other Handbooks for MAP, (BOWHILL and EDWARDS, 1983, 1984).

In this paper a short introduction to the radar technique with special emphasis to MST radars is given in section 2, followed by an outline of characteristics of atmospheric turbulence which gives rise to the MST echoes (section 3). In sections 4 and 5 the data acquisition and digital preprocessing as well as some fundamentals of data analyses are discussed. In section 6 techniques to measure velocities are treated and elucidated. Fundamentals of antenna designs are sketched in section 7 and a brief overview on the transmitter-receiver system and the radar control is presented in section 8. In section 9, empirical estimates of MST radar sensitivity, supported by a few typical results, are discussed. Finally, in section 10 a list of planned and operational MST radars as well as a short outline of expected future developments is added.

2. INTRODUCTION TO MST RADAR PRINCIPLES

(2.1) Pulse Operation

Let a pulsed electromagnetic wave be transmitted at the time T_1 (see Figure 1). The pulse duration of this radar signal shall be Δt_r. For simplification the pulse shape is assumed to be rectangular, but in real applications it may be a smoothed trapezoid or triangle or Gaussian shaped. In a nondispersive propagation medium the pulse travels with the speed of light C and reaches the range r_a after the time $t_1 = r_a/C$. A target at r_a can scatter or reflect the radar signal in some directions. A small fraction re-
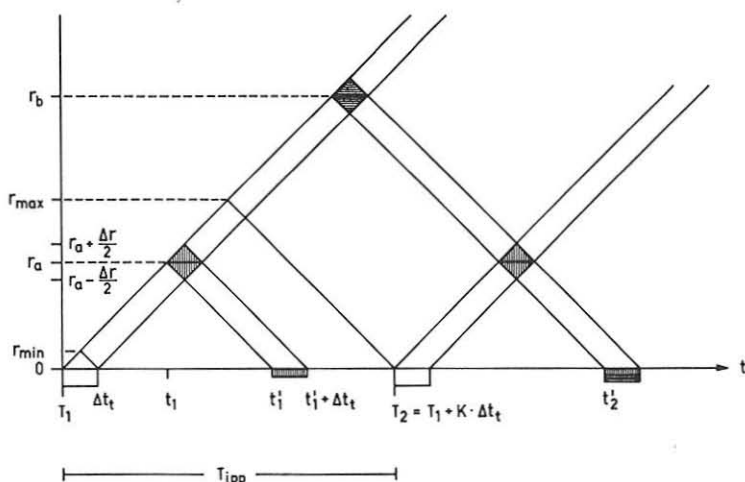
Figure 1. Range-time diagram.

turns to the location of the transmitter, where this radar echo will be received after the time $t_1' = 2t_1 = 2r_a/C$. This yields the basic relation $r = C \cdot t/2$, which allows determination of the range of any radar target by measuring the round-trip time t. This relation holds for monostatic radars (transmitter and receiver are at almost the same location). For bistatic radars (receiver separated from transmitter by a distance comparable to or larger than the ranges to the target) a modified expression has to be applied. Bistatic or multistatic operation will not be discussed here, since it is not too suitable for most MST radar applications.

Since the transmitted pulse has a finite duration Δt_t, its trailing edge will reach the range r_a at a time $t_1 + \Delta t_t$, and reach the receiver at $2t_1 + \Delta t_t$. If the radar target is a thin reflecting surface (called discrete, single or hard target), the shape of the received pulse is the replica of the transmitted pulse. Now assume that many scatterers fill all ranges along the radar beam. It is then obvious from Figure 1 that echoes from the range between $r_a - \Delta r/2$ and r_a reach the receiver simultaneously at t_1', and those from r_a to $r_a + \Delta r/2$ are received at $t_1' + \Delta t_t$. The pulse of duration Δt_t, thus, at one time illuminates a volume at r_a extended along a range $\Delta r = C \cdot \Delta t_t/2$. This is the range gate or range cell from which the radar echoes are received. Figure 1 shows that most of the echo power results from the range r_a, and minimum power is received from $r_a \pm \Delta r/2$. Thus, the resulting range weighting function of the single range gate centered around r_a is a triangle.

Because of the finite receiver bandwidth (usually matched to the bandwidth of the transmitted pulse) the receiver gate has a finite width in time. The receiver pulse response (given by about the inverse of the receiver bandwidth) distorts the received pulse by smoothing its leading and trailing edges and thus delaying it by about the receiver response time. The transmitter pulse length Δt_t, to which the receiver response time is matched, determines the range resolution Δr. The instrumental distortions of the radar pulse are not depicted in Figure 1 to maintain the clearness of survey, but they eventually have to be considered in the final analysis of the radar data.

In radar applications short pulses are normally transmitted periodically, i.e. the n-th pulse follows the (n-1)-th pulse after a specified time. For convenience it is set here to be a multiple (K>1) of Δt_t. This time ($T_n - T_{n-1}$) is called the interpulse period T_{IPP}. Its inverse is called the pulse repetition frequency $f_{PRF} = 1/T_{IPP}$. The off-on-ratio of the transmitter $T_{IPP}/\Delta t_t - 1$ determines approximately the range from which radar echoes can be unambiguously received (in units of range resolution). It is more customary, however, to use the ratio $d = \Delta t_t/T_{IPP}$, which is called duty cycle or duty factor. The transmitted power P_a, to be averaged over (more than) one interpulse period, is the product of the duty cycle and the transmitted pulse peak power P_p, i.e. $P_a = d \cdot P_p$.

Because in normal radar operations the pulse repetition frequency is kept constant, i.e. the transmitted pulse train is periodic, range-aliasing may occur. This ambiguity is depicted in Figure 1. At time t_2' an echo of pulse T_2 is received from range r_a, and an echo of pulse T_1' is received from range r_b. Of course higher order range-aliasing can occur from ranges $r_n = C \cdot (t + (n-1) \cdot T_{IPP})/2$. Because these echoes return from separate scatter volumes, the echo signals are uncorrelated but still their power accumulates in the same receiver range gate. If no special arrangements (e.g. pulse-coding or non-periodic T_{IPP}) are being made, the maximum unambiguous range is $r_{max} = C \cdot T_{IPP}/2$. The minimum range r_{min} obviously is given by the pulse duration Δt_t plus some instrumentally entailed transition time between transmission and reception.

2.2 Doppler Measurements

Assume that a bulk motion carries the scatterers or reflectors in the volume at range r. Because of the Doppler effect, the rate of change of phase ϕ of the returned signal is then $d\phi/dt = 4\pi/\lambda_o \cdot dr/dt$, where λ_o is the radar wavelength. When V' is the (radial) velocity in direction of the radar signal path, $V' = dr/dt$. The phase change $d\phi/dt$ is the angular Doppler frequency $\omega_D = 2\pi f_D$, which yields $f_D = -2V'/\lambda_o$. Since the radar signal is pulsed at a frequency f_{PRF}, i.e. the radar echo is sampled at a rate T_{IPP}, this yields the maximum Doppler frequency to be resolved by pulse-to-pulse analysis (Nyquist frequency): $f_{Dmax} = f_{PRF}/2 = 1/2T_{IPP}$. If coherent preintegration over 2N interpulse periods is applied (section 4.3), and the effective sampling time is increased to $2N \cdot T_{IPP}$ (section 4.3), and f_{Dmax} has to be replaced by $f_N = 1/4NT_{IPP}$. This corresponds to a maximum radial velocity $V'_{max} = \lambda_o \cdot f_{Dmax}/2$ and $V'_{max} = \lambda_o \cdot C/8r_{max}$. It turns out for MST radars that V'_{max}, as defined here, is much larger than any realistic velocity. We will readily recognize this, as well as deduce some obvious experimental advantages by estimating some typical numbers for the above defined radar parameters.

MST radars operate in the lower VHF band around 50 MHz, corresponding to wavelengths around 6 m. Since quasi-vertical antenna beam directions are used, ranges are roughly equal to altitudes. For MST radar observations of the middle atmosphere the range limits r_{max} are between 10 km and 100 km. This yields typical pulse repetition frequencies between 10 kHz and 1 kHz. Altitude resolutions from about 1 km down to at least 100 m are required to resolve typical vertical scales in the middle atmosphere. This corresponds to pulse lengths of about 1-10 μs. Thus, typical duty cycles are between about 10^{-1} and 10^{-3}.

Knowing that radial velocities with quasi-vertical radar beams do not exceed several 10 ms^{-1}, the Doppler frequency will barely exceed 10 Hz. Applying $f_{PRF} > 1$ kHz, the radar echo will be heavily oversampled, i.e. its phase and amplitude does vary little from pulse to pulse. This is called a coherent radar echo, in contrast to an incoherent radar echo which randomly

changes phase and amplitude from one pulse to the next. One does make efficient use of the characteristic coherency of MST radar echoes to improve the data acquisition procedures (section 4.3). The differences between coherent radars, used to study the lower and middle atmosphere, and incoherent (scatter) radars, used to study the ionosphere with the Thomson scatter technique, will be described in a different paper (ROTTGER, 1984).

2.3   Radar Equation for Scattering and Reflection

Assume that the radar echo power is due to volume scatter and the scatterers totally fill the radar beam. Then the mean received radar echo power P_s is given by the radar equation:

$$P_s = \frac{A \cdot P_a \cdot n \cdot \Delta r}{4\pi r^2} \qquad , \qquad (1a)$$

where A is the effective antenna area (see section 7.1) and n is the radar reflectivity (see section 3.2).

If the radar echo is due to reflection from a hard target, e.g. a large surface stratified perpendicular to the radar wave propagation, the received radar echo power P_r is given by:

$$P_r = \frac{P_a \cdot A^2 \cdot |\rho|^2}{\lambda_o^2 \cdot r^2} \qquad , \qquad (1b)$$

where ρ is the amplitude reflection coefficient of the surface.

As compared to radar echoes from single hard targets (e.g. airplanes), where the echo power is proportional to r^{-4}, P_s and P_r are proportional to r^{-2}. This is simply explained by the fact that the volume or the region illuminated by the radar beam, are not constant but increase with the square of the range. However, restricting assumptions have to be fulfilled here: The most essential are that the illuminated volume has to be totally filled with scatterers and a reflecting surface has to be larger than the first Fresnel zone $(D = \sqrt{r \lambda_o})$. If more than one reflecting surface is in the radar volume, $|\rho|^2$ has to be replaced by $R^2 = \Delta r \cdot (F\overline{M})^2$, where F is a calibration constant and M is the mean generalized refractive index gradient (e.g., HOCKING and ROTTGER, 1983).

If one assumes that the mechanisms of reflection and scattering are independent of each other, the total received echo power is

$$\overline{P} = P_r + P_s = P_a \cdot A \cdot \Delta r \cdot (C_r^2 + C_s^2) \cdot r^{-2} \qquad (2)$$

with the contribution due to reflection $C_r^2 = A \cdot (F\overline{M})^2/\lambda_o^2$, and the contribution due to scattering $C_s^2 = n/4\pi$. Both, C_r^2 and C_s^2, are dependent on properties of the reflecting and scattering media, evaluated at scales of half the radar wavelength λ_o and C_r^2 is additionally weighted by the instrumental parameters, antenna area A and the wavelength λ_o.

The theoretical treatments of these processes are based on the same formalism, namely the Born approximation, i.e. the amplitude of the scattered spherical wave is negligible compared to the amplitude of the incident plane wave. The three-dimensional statistical approach is based either on isotropic or on anisotropic volume scatter. The one-dimensional approach, treated in a statistical manner, describes Fresnel scatter which leads to similar results as for anisotropic volume scatter. The one-dimensional deterministic approach

describes the reflection process.

Following early observations with MST radars operating in the lower VHF-band, it is now accepted that scattering and reflection contribute to the radar echoes (GAGE and GREEN, 1978; ROTTGER and LIU, 1978; GAGE and BALSLEY, 1980; ROTTGER, 1980b), but fairly often the limits between these two processes are not readily determinable in an experiment. It was therefore proposed to use the term effective reflectivity $C^2 = C_r^2 + C_s^2$. Depending on whether C_r^2 or C_s^2 is dominating the signal (see ROTTGER, 1980b, and RASTOGI and ROTTGER, 1982, for more detailed discussions), the basic processes are either called reflection or scattering.

Different terms are used to specify these processes, namely specular, partial or Fresnel reflection if a single reflecting surface dominantly contributes to the radar echo, and Fresnel scatter if several statistically independent surfaces in the radar range gate produce the echoes. If it is found that the reflecting surfaces are rough and corrugated, then the term diffuse reflection is used. Also isotropic volume scatter is observed.

3. ATMOSPHERIC TURBULENCE CAUSING VHF RADAR ECHOES

3.1 Characteristics of Turbulence Spectra

The basic condition for any kind of scatter and reflection is the existence of spatial variations of the refractive index at scales of half the radar wavelength. These variations can be caused by atmospheric waves or by turbulence. The turbulence can be active (just generated) or fossil (remnants of active turbulence), and it is partly generated by waves. It is often a question of how to distinguish turbulence from waves, but one still can describe the energy cascade of atmospheric waves and turbulence by means of their spatial and temporal spectra. These items are not discussed here and more information can be found elsewhere, e.g. GAGE (1979), KLOSTERMEYER (1981), WOODMAN (1981), LARSEN (1983a), VANZANDT and VINCENT (1983), FRITTS et al. (1984).

The spatial spectrum of atmospheric turbulence is basically characterized by three subranges. In the buoyancy subrange, variations are controlled by buoyancy forces and the scales of the anisotropic eddies are typically larger than a scale L_B which is some ten to hundred meters in the middle atmosphere. Since these scales are larger than the wavelengths of VHF radars, they do not contribute to the scattering/reflection process. There is a transfer of energy from eddies in the buoyancy subrange through the inertial subrange to smaller eddies. The subrange transition is at L_o, which is also called the outer scale of the inertial subrange or the Kolmogorov macroscale. Turbulence at scales of several meters in the inertial subrange yields the VHF radar echoes. At the inner scale l_o (also called Kolmogorov microscale) another transition takes place into the viscous subrange. The turbulence, decaying down to scales in this range, is heavily damped by viscosity, and refractive index variations at scales smaller than l_o are so small that they do not contribute to detectable radar signal scattering.

The Kolmogorov microscale l_o changes strongly as a function of altitude, because it is inversely proportional to the density. It also depends on temperature and turbulence intensity which causes uncertainties about its mean value by many 10%. The mean l_o is less than some centimeters in the lower troposphere. It exponentially increases with height to several centimeters near the tropopause, to about a meter near the stratopause and to about ten meters near the mesopause. VHF radars with Bragg wavelengths λ_B (half the operating wavelength λ_o for backscatter) of about 3 m are therefore able to detect echoes from altitudes up to near the mesopause. Depending on the scat-

tering/reflection mechanism, they could observe the entire altitude region of the mesosphere, stratosphere and troposphere, which led to the introduction of the name of MST radars. In contrast, radars operating in the upper VHF band ($\lambda_B \approx 1$ m) and in the UHF band ($\lambda_B = 5$-50 cm) can detect echoes from turbulence only up to at most the middle stratosphere. Therefore, these are called ST radars. We do not include the incoherent or Thomson scatter radars in this nomenclature, which detect echoes from free electrons in the ionosphere, and thus can cover the altitude range of the mesosphere and the thermosphere (see MATHEWS, this handbook volume).

3.2 Turbulence Refractive Index Structure Constant

The reflected and scattered echo power is given by the component of the spatial spectrum of the variation of the refractive index n (also called refractivity here), whose wavelength is one half the radar wavelength. The reflected echo power depends on the shape n(r) of the refractive index discontinuity. In general, the steeper the gradient of the refractive index the larger the reflection coefficient; steps in the refractive index must not be much greater in spatial extent than a quarter wavelength, otherwise destructive interference occurs and will strongly reduce the reflection coefficient. It obviously follows that a longer-wavelength radar will see a larger reflection coefficient than a shorter-wavelength radar for a given profile of refractive index. Different models of refractive index profiles can yield the same reflection coefficient. This means that a deduction of refractive index variations from the reflected radar echo is generally ambiguous.

Let us assume that atmospheric turbulence mixes the refractive index profile and the associated gradients so that random irregularities of the refractive index result. The power which is scattered back then no longer results from reflection at a deterministic discontinuity, but from "reflection" at many disordered turbulent irregularities, which fill the volume defined by the pulse length and the antenna beam width (radar volume). Simply put, each irregularity has a single backscatter cross section. The total backscatter cross section per unit volume is the radar reflectivity η, which is related to the turbulent variations of the refractive index.

It is assumed that the turbulence is in the inertial subrange, which is given by the boundaries of the outer scale L_o, associated with macroscale turbulent eddies, and the inner scale l_o, where the turbulent irregularities start to be dissipated by viscosity. In the inertial subrange the turbulence is assumed to be homogeneous and isotropic. Under these assumptions the radar reflectivity is $\eta = 0.39 \cdot C_n^2 \cdot \lambda_o^{-1/3}$. The quantity C_n^2 is called the turbulence refractive index structure constant. It is given by

$$C_n^2 \sim \overline{\Delta n^2} \cdot L_o^{-2/3},$$

and is a measure of the outer scale L_o of the inertial subrange of the turbulence spectrum and the mean square variations $\overline{\Delta n^2}$ of the refractive index n (e.g. BATTAN, 1973).

Since turbulence parameters determine the structure constant C_n^2, which in turn determines the radar reflectivity η and the scattered power P_s, it appears likely that information on turbulence can be deduced from the radar observations. Qualitative information (e.g. morphology of turbulence layers and their strength, intermittency and persistency) is more definitely obtained than quantitative information (e.g. turbulence energy dissipation, diffusivity and viscosity, momentum transfer and wave breaking dynamics). The latter needs a very detailed treatment, since instrumental effects, separability of reflection and scattering, and further geophysical parameters (e.g. the mean gradient of refractive index) have to be considered. It is not in the scope of this

paper to discuss these in detail. They, for instance, were treated by VANZANDT et al. (1981), HOCKING (1983a,b) and others. We only would like to outline those items which are most general to the application of the MST radar technique: 1) some basic information on turbulence is needed to estimate the sensitivity of radars, 2) echoes from refractivity variations are used as tracers to monitor the structure of the atmosphere and to measure bulk and fluctuating velocities.

The MST radar observations have shown that the turbulence at scales of some meters is often anisotropic. This causes some uncertainties in the deduction of turbulence parameters, but on the other hand gains a substantial increase of echo power, i.e. signal detectability due to reflection from the anisotropic irregularities. Strong evidence is found that isotropic and anisotropic turbulence irregularities coexist. A model which was earlier proposed by BOLGIANO (1968) has recently found new interest and can explain some of the gradients of the background refractive index. In a turbulent layer vertical mixing tends to equalize mean gradients of the background refractive index. This consequently causes discontinuities at the top and the bottom of the layer. Since the turbulent layers have larger horizontal than vertical dimensions, these discontinuities of the refractive index have much larger horizontal than vertical dimensions, too. The discontinuities must often be as thin as several meters (in vertical extent) since they cause reflection of MST radar signals. They are also called stratifications, laminae or sheets. It is deduced from the observations that they are rough or corrugated because of the interaction with the turbulence layer. These sheets are often very persistent and it is assumed that they have a much longer lifetime than the originating turbulence layer, which one may call "fossil turbulence". Another idea was recently brought forward by VANZANDT and VINCENT (1983) who proposed that the anisotropic reflectivity is due to low frequency bouyancy waves. Both models still require experimental verification.

(3.3) Origin of Refractivity Changes

The refractive index variations discussed in the preceding paragraphs are directly related to variations of the atmospheric parameters: humidity, temperature, pressure ($\triangleq$ neutral density) and electron density. The mean refractive index in the lower and middle atmosphere is for VHF:

$$n = 1 + n_1' + n_2' + n_3',$$

where $n_1' = 3.7 \cdot 10^{-1} \cdot e/T^2,$

$n_2' = 77.6 \cdot 10^{-6} \cdot p/T,$

and $n_3' = -40.3 \cdot N_e/f_o^2,$

with e = partial pressure of water vapor (humidity) in mb,
      p = pressure in mb,
      T = temperature in K,
      N_e = number of free electrons per m^3,
      f_o = radar operating frequency in Hz.

n_1' is called the wet term, n_2' the dry term, and n_3' the ionospheric term.

All three terms are very small compared to 1; their contribution to the refractive index yields only less than a few parts per thousand change. The variations of n, which cause the scattering and reflection, are evidently produced by turbulence-induced variations of e, p, T and N_e. Although this is not too realistic, we assume for illustration that the average intensity of turbulence is constant with altitude. We then can estimate the relative contributions of e, p, T and N_e to the radar reflectivity, which is shown in

Figure 2 (after GAGE and BALSLEY, 1980). The wet term n_1' (humidity contribution) has to be considered only in the lowest few kilometers of the troposphere, whereas the dry term n_2' (temperature contribution) dominates up to the lower mesosphere. Contributions due to pressure fluctuations are normally negligible, but decreasing mean pressure with altitude has a strong influence on n_2'. For radars operating in the lower VHF band, the ionization of the D region determines the refractive index in the height region between about 60 km and 90 km. Except of the ionospheric term n_3', the refractive index contributions are nondispersive, i.e. frequency independent. Polarization and absorption effects can be neglected for VHF and UHF signals in the entire height range of the troposphere, stratosphere and mesosphere. At frequencies larger than about 50 MHz, the turbulence-induced scatter term will get very weak in the mesosphere (because of the viscous subrange limitations) and the incoherent scatter term will dominate the signal. For scales in the inertial subrange, clear-air turbulence in the mesosphere yields perturbations in the ionization (D-region irregularities) and it has to be stressed that these irregularities are not created by plasma instabilities. They are induced by turbulence in the neutral atmosphere because of the collisions between ionized and neutral molecules in this height region. The D-region irregularities are therefore replica of neutral air turbulence.

We could assume from Figure 2 that the entire altitude range up to 100 km could be monitored by VHF radars. However, apart from the inertial subrange limiting the VHF radar observations to altitudes below about 90 km (above this height occasionally only meteor echoes can be used), other limitations have to be considered. The refractivity contributions obviously change during varying atmospheric conditions. The electron density in the ionospheric D region (60-90 km) is directly dependent on the solar zenith angle, and consequently the radar reflectivity of the mesosphere is relatively high only during the daylight hours. Solar flares can also generate short enhancements of the D-region ionization. At high latitudes additional ionization is often generated even during the night by particle precipitation effects during auroral disturbances. However, the strong dependence of the electron density with height very seldom yields sufficiently strong reflectivity for echoes to be detected from heights below 60-65 km. Echoes from meteor-induced ionization can be used to measure velocities in the altitude range 90-100 km (AVERY et al., 1983). The stratospheric reflectivity is only determined by temperature vari-
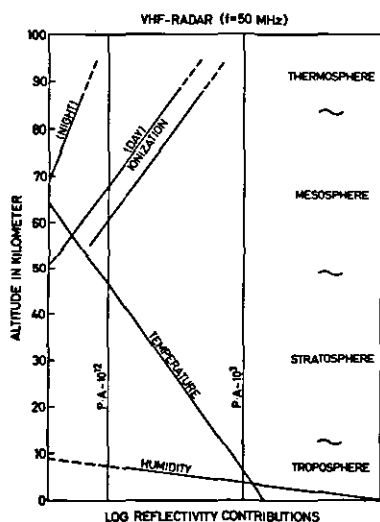

Figure 2. Altitude variation of relative reflectivity contributions for VHF radars (after GAGE and BALSLEY, 1980).

ations, and this mostly holds also for the reflectivity in the upper tropo-
sphere. However, during convective processes water vapor may be carried up to
the tropopause, which then yields a substantial increase of reflectivity (by
n_1') in the upper troposphere. In the tropics the reflectivity may even be
enhanced due to humidity contributions up to 15-18 km altitude.

The mean profile of reflectivity contributions shown in Figure 2 is not
only varying because the averages n_1', n_2' and n_3' are varying, but is
evidently modulated by the occurrence of turbulence and stable stratifications.
Higher reflectivities for instance occur in regions which are most likely
to be turbulent, e.g. jet streams in the upper troposphere and breaking of
tides and gravity waves in the mesosphere, as well as in regions of enhanced
static stability, e.g. the lower stratosphere.

As mentioned earlier, the radar signal strength depends on the product of
the instrumental parameters, average power and antenna area (aperture), and
the atmospheric parameter, effective reflectivity. The detectability of an
echo is a function of the signal-to-noise ratio, and the noise level at VHF
is given by the sky noise, which is constant when averaging over a day. The
sensitivity or the minimum detectable reflectivity of MST radars therefore is
directly proportional to the power-aperture product $P \cdot A$. For illustration,
typical low ($P \cdot A = 10^3$ Wm^2, e.g. early tests with the SOUSY-VHF-Radar
(ROTTGER et al., 1978)) and high (Jicamarca VHF radar (BALSLEY, 1978a; FUKAO
et al., 1979)) power-aperture products are inserted in Figure 2 from which the
observable height regions can be estimated approximately. These limits are
for vertical antenna beams, i.e. they are essentially determined by reflection
processes. The limits shift to the right for off-vertical beams when no re-
flection but only scattering is observed. These qualitative estimates should
be used as guidelines only, and more quantitative estimates are discussed in
section 9.

4. MST RADAR SIGNAL ACQUISITION AND PREPROCESSING

The simplified schematics of an MST radar system, shown in Figure 3, are
used for a basic explanation of the MST radar operation and data-acquisition
procedure. In an operational system many more components are used, but Figure
3 shows only those which are necessary for this explanation. For more details
see WOODMAN and GUILLEN (1974), ROTTGER and SCHMIDT (1979), SCHMIDT et al.,
(1979), CARTER et al. (1980), CLARK and CARTER (1980), SATO and WOODMAN (1980),
RASTOGI (1983), for instance.

4.1 Coherent Detection

An oscillator generates a signal s^o at the angular frequency $\omega_o = 2\pi f_o$
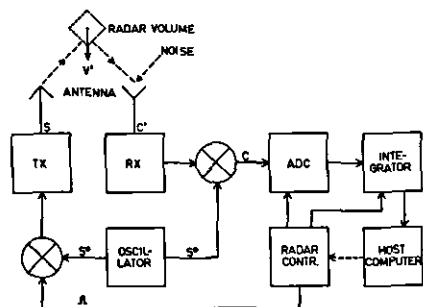where $f_o = C/\lambda_o$ is the center radar operation frequency. A pulse

Figure 3. Basic principle of an MST
radar system.

train, generated by the radar controller, imposes a modulation to this signal. After amplification in the transmitter (TX) the radar signal

$$s(t) = a(t) \exp(i(\omega_o t + \phi(t))) \tag{3}$$

is transmitted, where $a(t)$ determines an amplitude modulation (by the pulse train) and $\phi(t)$ corresponds to a phase modulation (for coding), and $i = \sqrt{-1}$. $a(t)$ and $\phi(t)$ are slowly varying as compared to $\omega_o t$. The radar signal is scattered/reflected from the radar volume and reaches the receiver (RX) via the same or a separate antenna. Additionally, noise (sky noise and interference) is received and adds to the radar echo. The band-limited echo signal s' plus noise r can be represented by

$$c'(t) = s'(t) + r(t) = a_1(t)\cos \omega_o t + ia_2(t)\sin \omega_o t,$$

where a_1 and a_2 are independent Gaussian variables in a pure scattering process, and are correlated in a reflection process. The uncorrelated noise contributes only uncertainties to these estimates. After linear amplification in the receiver, $c'(t)$ is coherently detected by multiplicative mixing with s^o. After low-pass or post-detection filtering (to eliminate high frequency components $2\omega_o$, which are generated during mixing), this yields

$$c(t) = a^+(t)\cos \psi^+(t) + ia^+(t)\sin \psi^+(t),$$

where $\quad a^+(t) = (a_1^2(t) + a_2^2(t))^{1/2}/2,$ $\qquad$ (4a)

$$\psi^+(t) = \arctan (a_2(t)/a_1(t)).$$

The phase $\psi^+(t) = \omega_D t - \psi'(t)$ is given by the Doppler frequency $\omega_D = -4\pi V'/\lambda_o$, which is due to the bulk motion V' of the scatterers. The time variable phase $\psi'(t) = \psi(t) + \phi(t)$, where $\psi(t)$ is caused by the fluctuations of the scatterers/reflectors in the radar volume. The amplitude $a^+(t)$ is a measure of the reflectivity. These latter statements are only valid if the noise contributions are separated from the signal. The coherently detected complex signal (+ noise) can be expressed in the form

$$c(t) = x(t) + iy(t), \tag{4b}$$

where the real part $x(t) = a^+(t) \cos\psi^+(t)$ is called the in-phase component, and the imaginary part $y(t) = a^+(t)\sin\psi^+(t)$ is called the quadrature component. Both components, x and y, are called the quadrature components.

The Fourier transform of $c(t)$ is $\tilde{A}(\omega) = \int c(t)\exp(-i\omega t)dt = \tilde{x}(\omega) + i\tilde{y}(\omega)$, which yields the power spectrum $P(\omega) = \tilde{x}^2 + \tilde{y}^2$. The measured $P(\omega)$ is the convolution of the spectrum of the refractivity fluctuations in the radar volume with the spectrum $P_T(\omega)$ of the transmitted wave form, multiplied by the bandpass characteristics $P_R(\omega)$ of the receiver. Since $P(\omega)$ is much narrower than the envelope of $P_T(\omega)$ and $P_R(\omega)$ in MST radar investigations, these instrumental effects can mostly be disregarded.

4.2 Digital Sampling

In the analogue-digital converter (ADC) the signal $c(t)$ is sampled at discrete time intervals $t_k = k \cdot \Delta t_s$. This is illustrated in Figure 4, which shows the amplitude variations of one of the quadrature components as they can be monitored with an oscilloscope connected to the output of the quadrature detector (after postdetection filtering). The sequences $n = 1,2,...$ can be assumed to represent successive oscilloscope beam deflections, triggered by the leading slope of the transmitter pulse. This pulse is strongly at-
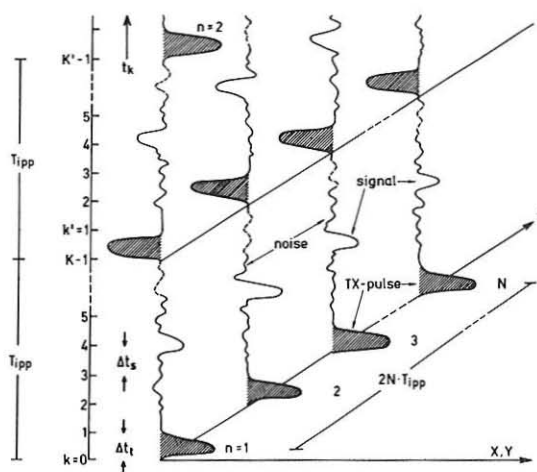
Figure 4.   Range-time-amplitude diagram.

tenuated by receiver gating. It is flipped in phase ϕ by 180° from one pulse
to the next (change from positive to negative amplitude), for reasons explained
later.   Corresponding to the transmitter-pulse phase ϕ, the sign of the signal
also changes from one to the next interpulse period. For convenience the
signal phase is shown here to be similar to the transmitter-pulse phase. The
signal phase can take any values, however, depending on the length of the phase
path from the transmitter to the radar volume.

On the vertical axis of Figure 4 the time t_k is given, which directly
can be converted to range by means of the definitions of Figure 1. The sampl-
ing time interval Δt_s should be equal to the radar pulse length, since this
yields an optimum matching to the range gate width or resolution Δr. The
signal and the noise had passed the receiver and postdetection filters and are
therefore bandlimited. The response time of the receiver, which is appoxi-
mately inversely proportional to the filter bandwidth, should also be equal to
Δt_s, respectively Δt_t. The subscript k of t_k is the serial number of
the range gate with $k = 0, 1, \ldots, K-1$, where $k = 0$ corresponds to the begin-
ning of the transmitter pulse. Since the quadrature components are digitally
sampled, we can write (4b) in the form $c_k = x_k + iy_k$. We also call c_k
the complex raw data samples.

The number of sampling time steps between successive radar transmitter
pulses is K, which is also the number of sampled range gates. The interpulse
period is $T_{ipp} = K \cdot \Delta t_s$. The cycle $k = 0, \ldots, K-1$ is repeated once with
$k' = 0, \ldots, K'-1$ with a reversed phase of the transmitter pulse. Both
cycles, denoted by k and k' (with $k' \hat{=} k$ and $K' \hat{=} K$), determine one radar
cycle $2T_{IPP}$. The serial number of radar cycles is given by $n = 1, \ldots, N$.
One radar burst is determined by N radar cycles, which last for $t_i = 2N \cdot T_{IPP}$.
The generation of all pulse trains, needed to control these cycles of the
transmitter and receiver-ADC-integrator system, is done in the radar controller
(Figure 3). It is loaded by the host computer which is also used for further
data processing.

Let us look at the time development of the signal + noise amplitude in
the k-th range gate, which is sketched in Figure 5. In this example the digi-
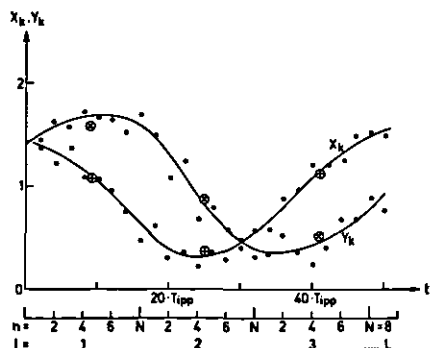tally sampled quadrature components x_k and y_k consist of a quasi-harmonic

Figure 5. Time display of quadrature components in a fixed range gate.

oscillation (of period of about $50T_{IPP}$, regarded as signal) with superimposed random fluctuations (regarded as noise). The signal + noise is lifted by a bias up to the mean amplitude 1. The bias can be different for the x- and y-components but for convenience the bias is chosen here to be equal for both. The digital samples (dots and circles) are taken every second interpulse period (k). Samples of the intermediate pulse period (k') lie between the de-picted samples. They would show an oscillation shifted by 180°, because of the transmitter phase flip from k to k'.

It is evident that the signal is oversampled, i.e. many more samples are taken than would be necessary to resolve the amplitude and phase of the har-monic oscillation. However, the noise is undersampled, since its time scale of fluctuation is much faster than the sampling rate given by T_{IPP}. The time scale of the noise fluctuations is proportional to the receiver response time ($\sim\Delta t_s$), and the time scale of the signal variations is given by the typical time scales τ_ϵ and τ_ϵ' of the scattering/reflection process. These latter times are inversely proportional to the Doppler frequency f_D and the average statistical phase changes $d\psi/dt$. To give some typical numbers:

$$\Delta t_s = 10^{-6}s \text{ (for 150 m range resolution)},$$

$$T_{IPP} = 10^{-3}s \text{ } (f_{PRF} = 1kHz),$$

$$\tau_\epsilon' = 1/f_D > 10^{-1}s \text{ } (f_D = 10 \text{ Hz}),$$

$$\tau_\epsilon = (d\psi/dt)^{-1} \sim 10^{-1}s \text{ (see section 5.2).}$$

We will see later that τ_ϵ and τ_ϵ' can be more appropriately expressed by the autocorrelation function analysis. It is evident, that $\Delta t_s \ll T_{IPP} \ll \tau_\epsilon$, τ_ϵ', which proves that the time scales of noise Δt_s are undersampled and the time scales of signal τ_ϵ are oversampled.

The undersampling of noise, which has a bandwidth of the receiver post-detection filters, cannot be avoided because samples at a specified range gate cannot be taken more often than every interpulse period T_{IPP}. This under-sampling means that all noise power (received within the receiver bandwidth) folds into the sampled spectrum which is limited by the Nyquist frequency $\pm 1/2 \ T_{IPP}$. Since the noise samples are totally independent, because $\Delta t_s \ll T_{IPP}$, the undersampling does not cause problems. Because τ_ϵ, $\tau_\epsilon' \gg T_{IPP}$, the raw data signal samples are not independent, i.e. coherent.

4.3 Preintegration and Processing

The spectrum which one would obtain with the sampling rate T_{IPP} is very

wide and mostly contains high frequency noise power. The signal power is confined to relatively low frequencies only ($f = 1/\tau_\varepsilon \ll 1/2T_{IPP}$). It is evident therefore that low pass filtering, done before the spectrum analysis, will not change the signal characteristics but eliminates high frequency noise contributions. The simplest form of low pass filtering is just the complex addition of the signal + noise samples over an interval $t_i \ll \tau_\varepsilon$. This is illustrated in Figure 5, where the larger circles indicate the averages over $N = 8$ samples in each case of this example. A readily noticed effect of this averaging is the reduction of the number of total samples by a factor $1/N$.

Since the noise (r) and the signal (s) are independent of each other, their quadrature components add to $c_{kn} = c_{kn}{}^r + c_{kn}{}^s$, where $c_{kn}{}^r = x_{kn} + iy_{kn}{}^r$ are the quadrature components of the noise and $c_{kn}{}^s$ those of the signal. The bias shown in Figure 5 can be due to an instrumental offset $c_{kn}{}^i$, or due to radar clutter $c_{kn}{}^c$ (clutter = echo from a fixed target). All these contributions are additive:

$$c_{kn} = c_{kn}{}^r + c_{kn}{}^s + c_{kn}{}^i + c_{kn}{}^c .$$

We have to take into account that each radar cycle yields two samples per range gate, namely k and k'. The samples of signal and clutter are shifted by 180° from k to k', because the phase $\phi(t)$ of the transmitter was flipped by 180° (change $\phi(t)$ by 180° in equations (3) and (4)). This can be accounted for by changing the sign of $c_{k'n}$ when averaging

$$\overline{c}_{k,k'} = \frac{1}{N} \sum_{n=1}^{N} (c_{kn} - c_{k'n})$$

$$= \frac{1}{N} \sum_{n=1}^{N} (c_{kn}{}^r - c_{k'n}{}^r + c_{kn}{}^s - c_{k'n}{}^s + c_{kn}{}^i - c_{k'n}{}^i + c_{kn}{}^c - c_{k'n}{}^c).$$

Since for the instrumental bias $c_{kn}{}^i = c_{k'n}{}^i$, it is eliminated by averaging. This is called instrumental-dc elimination (dc = direct current, better to say: constant voltage contribution).

Because of the transmitter phase flip: $c_{kn}{}^s = - c_{k'n}{}^s$ and $c_{kn}{}^c = - c_{k'n}{}^c$. Since the noise is independent from one to the next interpulse period, a change in sign of $c_{kn}{}^r$ does not change its statistical properties. We thus obtain, since $k' \hat{=} k$:

$$\overline{c}_k = \frac{2}{N} \sum_{n=1}^{N} (c_{kn}{}^r + c_{kn}{}^s + c_{kn}{}^c). \tag{5}$$

This averaging, commonly called "preintegration", has become a standard process in MST radar operations. It yields preintegrated data samples $\overline{c}_{kl}$ (l=1, ...L) at the time $t_l = 2 \cdot l \cdot N \cdot K \cdot \Delta t_s$ for the preintegration period $t_i = 2 \cdot N \cdot K \cdot \Delta t_s = 2 \cdot N \cdot T_{ipp}$.

If the summation in (5) extends over a time period $t_i = 2 \cdot N \cdot T_{IPP}$, which is much longer than Δt_s, the high-frequency noise contribution vanishes. For an integration period $t_L = 2 \cdot L \cdot N \cdot T_{IPP}$ (L>>N), which is much longer than τ_ε, the signal contribution approaches zero since it is slowly fading in amplitude and phase. Only the clutter contribution

$$\overline{\overline{c_k{}^c}} = 1/L \sum_{L=1}^{L} \overline{c}_{kl}$$

remains, since it is constant in amplitude and phase. This can be used to eliminate the clutter component by means of a digital high pass filter opera-

tion: $\tilde{c}_{kI} = \overline{c_{kI}} - \overline{c_k}^c$. This operation is called *clutter-dc elimination*, and is done after the preintegration.

For the preintegration given by (5), the number N of added samples has to be selected carefully. It is evident from Figure 5 that the integration period has to be much shorter than the typical time scale of signal variations. Suppose the integration period would extend over a full cycle of the signal oscillation in Figure 5. The integration then would yield $\overline{x}=0$ and $\overline{y}=0$. As can be seen from the displacement of the average values (crossed circles), the signal amplitude will be generally attenuated due to the preintegration. In the spectral domain the amplitude attenuation factor F is given by $F(f)=\sin(\pi f N T_{IPP})/\sin(\pi f T_{IPP})$. This effect of attenuation is also called the combfilter effect (SCHMIDT et al., 1979). The sin Nx/sin x-function arises because a non-tapered, rectangular window is used. The advantage is that the processing can be made very efficient since no multiplication of the raw data series with a weighting function is necessary. As long as care is taken that the first zero crossing of F(f), which is at $f_N=1/2NT_{IPP}$, is at a reasonably higher frequency than typical signal frequencies, F(f) can be used to correct the signal spectrum. After preintegration the maximum resolvable frequency f_{Dmax} (in section 2.2) has to be replaced by f_N.

It is also necessary that the real part x and the imaginary part y of the detected signal are correctly in quadrature (orthogonal). Ideally they must be phase-shifted by exactly 90° and must have equal amplitudes on the average. Accuracies of less than a few degrees phase difference and less than a few percent amplitude difference are tolerable, however, and can be obtained by proper hardware adjustment. If these accuracies cannot be reached, a correction of the quadrature components can be applied within some limits (e.g., RUSTER and WOODMAN, 1978).

The preintegration of the quadrature components, formulated by (5), is normally done in a digital preprocessor, called adder or integrator (see Figure 3). Since this preintegration is a low pass filter process it can be done also in an analogue filter, such as the clutter elimination which is a high pass filter operation. Obviously the digital processing is much more versatile and flexible. Both analogue and digital preintegration reduce the number of data samples by some orders of magnitude, and compress the huge data flow from the ADCs to make it manageable for the host computer. This is the evident advantage of this preintegration process. It is often thought that this process also increases the signal-to-noise ratio since the voltages of the coherent signal but the power of the incoherent noise add. This leads to an improvement of the signal-to-noise (power) ratio by the factor N, since the noise bandwidth is changed by the *factor* N^{-1}. *If one defines* the noise in such a way that its bandwidth is equal to the signal bandwidth, the preintegration process only reduces the wide noise bandwidth to the acceptable limit close to the signal bandwidth, and the signal-to-noise ratio does not change by this preintegration. The advantage of this preintegration process is still the essential reduction of the number of raw data samples, without giving away information on the signal.

4.4 Coding

Another preprocessing step which is generally similar to preintegration is the decoding procedure (described in detail in the paper by WOODMAN, SULZER and FARLEY in this Handbook). The principal reason for applying coding/decoding (pulse compression) is to achieve a maximum average power at optimum resolution and maximum unambiguous range. A fairly low duty cycle of 0.1% and hence a low sensitivity would for instance result from using a single pulse of 1 µs duration (Δr = 150 m) and an interpulse period of 1000 µs (r_{max} =

150 km). Increasing the pulse length to 32 μs would increase the duty cycle to 3.2%, but deteriorate the range resolution to 4.8 km. The range resolution of 150 m can still be achieved by phase-coding the transmitter pulse in time (lag) increments of, say 1 μs. The decoding has to be done by cross correlating the received complex time series c_k with the transmitted code s_m:

$$c_{k*}' = \sum_{m=0}^{M-1} c_{k*+m} \cdot s_m$$

where M is the length of the code and k* corresponds to the range gate k at lag zero where the correlation function c_{k*}' has a maximum. Thus, the decoding is nothing else but an integration over several range gate samples, which are multiplied by the weighting factor s_m.

The simplest and most versatile phase coding scheme is the binary code, where the phase is flipped between the two states -1(=0°) and +1(=180°). Since here the weights s_m are +1 or -1, the multiplication needed in the decoding process reduces to an addition/subtraction operation. For this reason the decoding can be done in a preprocessor similar the the hardware pre-integrator described earlier, or the integrator/decoder can even be one unit. An advantage in MST radar applications is also that the decoding/integration processes are interchangeable, which reduces the number of operations by about two orders of magnitude (e.g., WOODMAN et al., 1980).

The best codes for radar applications obviously are those where sidelobes of the correlation function at k* ≠ k are minimum. Reasonable sidelobe suppression is gained with Barker codes, where the phases are flipped in a sequence (e.g., ++++--++-+-+ for the 13-bit Barker code). The correlation function is 13,0,1,0,1,0,1,0,1,0,1,0,1. The best sidelobe suppression, achieveable with a Barker code, is M^{-1}, and the sidelobes extend out to $M \cdot \Delta t_s$. Since the codes cannot be infinitely long, because the minimum range is extended with the length of the code. the sidelobe suppression of Barker codes is limited. The reason is that these sidelobes contain power from other range gates k* ≠ k. which cause ambiguities.

A better sidelobe suppression can be gained by application of quasi-random codes, which however need decoding before the preintegration. The most suitable codes for MST radar work are the complementary codes (e.g., SCHMIDT et al., 1979), which theoretically have no sidelobes. A complementary code consists of a pair of two code sequences s_1 and s_2. These have the property that their correlation function sidelobes are exactly equal. but opposite in sign. Normally both code sequences are transmitted at one and the next interpulse period. and the range samples of these two periods are preintegrated and decoded separately. The coherent addition of the decoded sequences then yields the total elimination of the sidelobes (k* ≠ k) if the signal is coherent from one to the next interpulse period. The zero lag value k* = k contains the total signal amplitude. As an example: s_1 = ++-++-+, s_2 = +++---+-, yield the correlation functions c_1 = 8,-1, 0,3,0,1,0,1 and c_2 = 8,1.0,-3.0,-1,0,-1 and the addition $c_1 + c_2$ = 16.0, 0,0,0,0,0,0. The peak value at k* = k is 2M. There is still the shortcoming of a long code, that it extends the shortest observable range. This can be overcome by transmitting a sequence of a complementary code and a short single pulse in one radar cycle.

5. DATA ANALYSIS

5.1 Correlation and Spectrum Analysis

After appropriate preintegration one still has to deal with the noise which is remaining within the bandwidth given by the signal. The elimination of the noise and further signal processing can be done in the time domain by

covariance (correlation) analysis or in the frequency domain by spectrum analysis. Since the covariance function and the power spectrum are Fourier transforms of each other, both contain the same relevant information. Depending on the purpose and the feasibility of the analysis, either of both approaches, is used in practice (RASTOGI and WOODMAN, 1974; WOODMAN and GUILLEN, 1974; ROTTGER and SCHMIDT, 1979; SCHMIDT et al., 1979; CARTER et al., 1980; CLARK and CARTER, 1980; SATO and WOODMAN, 1980; RASTOGI, 1983).

The complex autocovariance function of the quadrature components $\bar{c}(t)$ is

$$\rho(\tau) = \int \bar{c}(t) \cdot \bar{c}^*(t + \tau) dt,$$

where τ is the temporal displacement, and the * denotes the complex conjugate. In digital form

$$R_k(\tau_j) = \frac{1}{L-j} \sum_{l=1}^{L-j} \bar{c}_{kl} \bar{c}_{kl+j}^* = \overline{\bar{c}_{kl} \bar{c}_{kl+j}^*}, \quad j=0,\ldots,J-1, \quad J \leq L,$$

where j is the lag parameter defining the lag $\tau_j = j \cdot t_i$. For a fixed range k, $\bar{c}_l = \bar{x}_l + i\bar{y}_l$, and the covariance function becomes

$$R(\tau_j) = (\overline{x_l x_{l+j} + y_l y_{l+j}}) + i(\overline{x_{l+j} y_l - x_l y_{l+j}})$$

$$= R_r(\tau_j) + i R_i(\tau_j),$$

$$\text{or } R(\tau_j) = |R(\tau_j)| \exp(i\phi(\tau_j))$$

$$\text{with } |R| = (R_r^2 + R_i^2)^{1/2}, \quad \phi = \arctan(R_i/R_r).$$

In radar applications the term correlation function is often used for $R(\tau_j)$ instead of covariance function. The correct definition of the correlation function $\rho(\tau_j)$ is given by the normalized covariance function: $\rho(\tau_j) = R(\tau)/|R(0)|$. As well as the autocorrelation function of the series $\bar{c}(t)$ we also compute cross-correlation functions for two different series, $\bar{c}_1(t)$ and $\bar{c}_2(t)$, in the spaced antenna applications.

The *power spectrum is the Fourier transform of* $\rho(\tau)$, weighted by $W(\tau)$:
$$P(\omega) = \int W(\tau) \cdot \rho(\tau) \cdot \exp(-i\omega\tau) d\tau$$
or in digital form

$$P_k(\omega_m) = \frac{1}{J} \sum_{j=0}^{J-1} W_j \cdot R_{kj} \cdot \exp(-i\omega_m \cdot j \cdot t_i),$$

where W_j is an arbitrary weighting function (e.g., $W_j = J^{-1}$ for all j, etc.) and $\omega_m = m \cdot \pi/J \cdot t_i$; $m = 0,\ldots, J$; $t_i = 2 \cdot N \cdot T_{ipp}$.

The first three moments m_0, m_1 and m_2 of the power spectrum yield the total signal power $\bar{P}$, the mean $\bar{\omega}_D$ of the spectrum, and the width $\bar{\omega}_w$ of the spectrum:

$$\bar{P} = m_0 = \int P(\omega) d\omega,$$

$$\bar{\omega}_D = m_1/m_0, \quad m_1 = \int \omega P(\omega) d\omega, \tag{6}$$

$$\bar{\omega}_w = (m_2/m_0 - (m_1/m_0)^2)^{1/2}, \quad m_2 = \int \omega^2 P(\omega) d\omega.$$

Moments can also be deduced from the covariance function, which then yield the signal power, Doppler shift and spectral width, if the spectrum is symmetrical (WOODMAN and GULLEN, 1974):

$$\bar{P} = |R(o)| = R_r(o),$$

$$\overline{\omega}_D = d\phi(o)/dt = \phi(\tau_1)/\tau_1,$$

$$\omega_w^2 = 2(1-|R(\tau_1)| \; / \; |R(o)|)/\tau_1^2. \tag{7}$$

The power $\overline{P}$ is a measure for the effective reflectivity, the Doppler frequency $\overline{\omega}_D$ determines the mean velocity, and the spectral width $\overline{\omega}_w$ gives, after correction (e.g. HOCKING, 1983a), an estimate of the turbulent velocity fluctuations and the turbulence dissipation rate.

5.2 Parameter Estimation

Before calculating the parameters $\overline{P}$, $\overline{\omega}_D$ and $\overline{\omega}_w$ the noise contributions have to be eliminated. The noise level can be estimated from a range gate where no signal is detectable, or from the same range gate when the transmitter is switched off. However, if the signal bandwidth is sufficiently narrower than the noise bandwidth the high frequency parts of the spectrum or the zero-lag value of $R_r(o)$ can be used to obtain the noise estimate. The latter estimate requires the interpolation of the signal covariance function from higher order lags to the zero-lag and the subtraction of the interpolated value from $R_r(o)$ to obtain the noise estimate.

The advantage of the correlation analysis is, that the power, the Doppler frequency and the signal coherency can be read directly from the autocorrelation function. The power $\overline{P}$ is the interpolated value of the real part of $R(\tau)$ at zero lag. If the imaginary part of $R(\tau > 0)$ is zero, $\phi = const$, $d\phi/dt = 0$, and consequently $\overline{\omega}_D = V' = 0$. The amplitude $|R(\tau)|$, which is equal to $R_r(\tau)$ for $\overline{\omega}_D = 0$, determines the persistency of the signal. The time lag at which $|R(\tau)| = 0.5 \cdot |R(o)|$ is called the statistical coherence time τ_ϵ, which is inversely proportional to the fluctuations of the radar signal. The deterministic coherence time τ_ϵ' is inversely proportional to the Doppler frequency $\overline{\omega}_D$ (given by the radial velocity). It is thus inversely proportional to $d\phi/dt$, which is essentially given by the imaginary part of the correlation function at small lags.

The uncertainty of the signal parameters depends on the number $M \cdot L$ of averaged samples and on the signal-to-noise ratio $SNR = \overline{P}/P_N$, wher P_N is the noise power within a bandwidth equal to the signal bandwidth. Averaging over M samples of the correlation function or the spectra, which are computed from L preintegrated samples, is called postintegration. The fractional uncertainty of the power estimate is

$$\epsilon = \Delta\overline{P}/\overline{P} = (1 + P_N/\overline{P}) \cdot (M \cdot L)^{-1/2}.$$

Say, the preintegration time was chosen to be $t_i = 1/3s$, corresponding to a maximum noise (= signal) bandwidth of 3 Hz. Then a one-minute postintegration corresponds to $M \cdot L = 180$. For $SNR = 1$, this yields $\epsilon = 0.15$. To obtain the same uncertainty for $SNR = 0.1$, the postintegration time t_p has to be 30 min, and for $SNR = 0.01$, $t_p = 42h$. The uncertainties of $\overline{\omega}_D$ and $\overline{\omega}_w$ depend in a similar way on $(M \cdot L)^{-1/2}$, e.g. WOODMAN and HAGFORS (1969). However, it is questionable if postintegration times of more than several hours are reasonable, since the signal parameters are not stationary over such a long time period. These constraints of signal-to-noise ratio and signal stationarity place a definite limit on the sensitivity of radars (see section 9). Obviously, for $SNR \ll 1$ an increase of the transmitted power by a factor Q would reduce the postintegration time by Q^{-2} to obtain the same uncertainty ϵ. This is the essential point one has to bear in mind when discussing the sensitivity of radars; it is not so much the marginal increase of maximum observable altitude by increased power but the improvement in time resolution.

The three parameters $\overline{P}$, $\overline{\omega}_D$ and $\overline{\omega}_w$ fully define a power spectrum with Gaussian shape

$$P(\omega) = \frac{\overline{P}}{\sqrt{2\pi}\,\overline{\omega}_w} \cdot \exp\left(- \frac{(\omega - \overline{\omega}_D)^2}{2\overline{\omega}_w{}^2}\right),$$

which is assumed to be a good approximation for many conditions. These parameters yield information of the radar reflectivity ($\overline{P}$), turbulence intensity ($\overline{P}, \overline{\omega}_w$), mean radial velocity ($\overline{\omega}_D$) as well as the mean velocity fluctuations ($\overline{\omega}_w$) and the stability ($\overline{P}, \tau_\varepsilon$) and persistency ($\overline{P}, \tau_\varepsilon$) of layered structures in the radar volume. The parameters can be deduced either directly by applying equations (6) or (7), by parameter fitting or iteration procedures. Special circumstances, e.g. signal aliasing, variable interference, fading ground clutter, ocean clutter echoes or the simultaneously occurring echoes due to reflection and scattering, demand the application of special methods. These are, for instance, discussed by ROTTGER (1980b), SATO and WOODMAN (1980), RASTOGI and HOLT (1981), RASTOGI and ROTTGER (1982), HOCKING (1983a,b).

6. MEASUREMENTS OF VELOCITIES

Let us first discuss the measurements of the mean velocity since these are used to determine wind velocities in the atmosphere (e.g. WILSON and MILLER, 1972; GAGE and BALSLEY, 1978). There are basically two methods to measure velocities with MST radars: One method uses a narrow radar beam swung into various directions and measures the Doppler frequency of echoes scattered from irregularities. This method is the Doppler-beam-swinging (DBS) method. Another method uses three or more spaced antennas. and the received signals are cross-correlated to determine the drift speed of the scattering/reflecting irregularities (BRIGGS, 1977; ROTTGER and VINCENT, 1978; VINCENT and ROTTGER, 1980; ROTTGER. 1981b). This method is the spaced-antenna-drifts (SAD) method. More often they are just called the Doppler and the spaced antenna (SA) methods. Both methods are schematically sketched in Figure 6. Their physical mechanisms are principally similar, as shown by BRIGGS (1980), but in praxis the technical implementation may favor one of these methods.
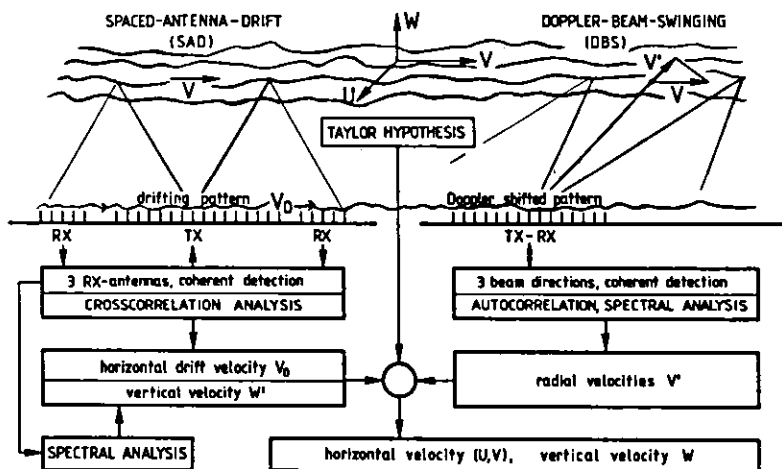
3-DIM VELOCITY MEASUREMENTS WITH VHF-RADAR

Figure 6. Principle of three-dimensional velocity measurements with MST VHF radars (from ROTTGER. 1981b).

6.1 Spaced Antenna Method

The spaced antenna method measures the temporal and spatial variations of
the field pattern of the radar echoes with vertically beamed antennas. As the
reflecting or scattering irregularities move through the beam they produce a
moving diffraction pattern on the ground which is sampled at three or more
spaced receiving antennas. By cross-correlating the received signals it is
possible to measure the time delays τ_a, between the receiving antennas and
hence the apparent velocity V_a. This can be corrected for random changes to
give the so-called true velocity V_t (see BRIGGS, 1977, and this Handbook for
more details). The vertical velocity component is estimated from the complex
autocorrelation or the spectral analysis.

When deducing the wind velocities from spaced antenna measurements, one
assumes that the drifting irregularities are carried with the wind (Taylor
hypothesis), which has been confirmed by comparing drift data with in situ
measurements. ROTTGER and CZECHOWSKY (1980) first presented results of simul-
taneous measurements with the spaced antenna and the Doppler method, which they
also compared with parallel aircraft and radiosonde measurements. In Figures
7a and 7b, examples of their comparisons with the spaced antenna method are
shown, where $|u| = (U^2+V^2)^{1/2}$ is the speed and $\alpha = \arctan(-U/-V)$ is the
direction of the wind, U and V are the zonal and meridional components. The
aircraft measurements (a) were in a few kilometers distance from the radar,
whereas the radiosonde measurements (b) were about 100 km distance north of the
radar. The aircraft and radar data are in excellent agreement, and the radio-
sonde and radar data are in good agreement. Because the latter measurements
were separated by more than 100 km, where the wind variability already cause
considerable differences, the similarity of both these wind profiles is con-
vincing. Of course, such comparisons have to be confined to the troposphere
and lower stratosphere, but their very reasonable outcome gains confidence that
wind measurments with this method are correct in the whole middle atmosphere
(papers by BRIGGS, FRASER and SCHMINDER in this Handbook). The comparisons of
ROTTGER and CZECHOWSKY (1980) showed also that the spaced antenna measurements
are evidently competitive with Doppler measurements.

Since its introduction to VHF radars, it appeared that the spaced antenna
method has advantages and disadvantages compared with other methods for measur-
ing winds. Because of the aspect sensitivity, it was argued that the spaced
antenna method is most suitable since it uses vertical beam antennas and one
need not worry about sidelobe effects. Obviously an increase of the signal-to-
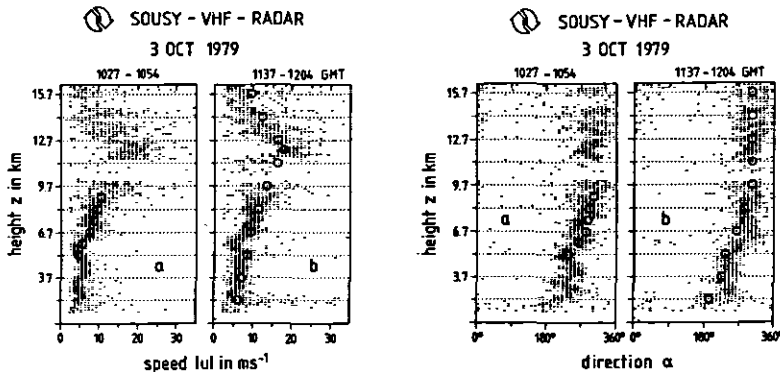

Figure 7. Wind speed and direction measured with spaced antenna
method (stippled distributions), aircraft (circles a) and
radiosonde (circles b).

noise ratio is gained and the fixed vertical beam positions are technically
much easier to handle than steerable antenna beams. However, also the spaced
antenna methods needs signals which are scattered and diffusively reflected
from off-vertical directions (BRIGGS, 1980). It is also evident that the field
pattern on the ground depends on the degree of horizontal coherency of the re-
flecting structures as well as on the width of the transmitter and receiver an-
tenna beams; i.e. a large structure coherency and large antenna apertures
(narrow beams) cause large correlation distances of the ground field pattern.
This places limits on the antenna size and separation. If, for instance, the
correlation distance is much larger than the antenna separation, a very broad
cross-correlation function ρ_a will result, which increases the error in de-
termining the lag τ_a. The optimum antenna spacings and sizes can be pre-
scribed only if correlation times and distances as well as horizontal velocities
are known. The best spacing of the receiving antennas would probably be such
that the field pattern drifts from one antenna to the other in not less than
some 10% of the correlation time τ_c. Of course, the sampling rate of the sig-
nal has to be short enough to resolve the correlation time. Observed average
values of τ_c are mostly around one second, although much higher and much lower
values can occur. Typical horizontal velocities are between some ms^{-1} to some
ten ms^{-1} with average values around about 20 ms^{-1}. This yields an optimum
spacing of several tens of meters and places an upper limit to the antenna
sizes, which in turn yields an upper limit of the sensitivity of a spaced an-
tenna system. However, for high signal-to-noise ratios, which normally charac-
terize signals from the troposphere and lower stratosphere, this antenna size
limit does not impair the application of the spaced antenna method. It rather
supports its application in small and cost-effective systems.

Of course, many more details, such as analyses procedures and data selec-
tion criteria, have to be considered when planning, designing and operating a
spaced antenna system, and the reader is referred to the comprehensive papers
by BRIGGS (1977, 1980), VINCENT and ROTTGER (1980), ROTTGER (1981b), HOCKING
(1983c), ROYRVIK (1983a). It was recently accepted that an MST radar spaced
antenna system can also be used in the interferometer mode, which yields ad-
ditional useful information on spatial structures of waves and turbulence. The
application of this technique is outlined in section 6.3.

6.2 Doppler Methods

In contrast to the spaced antenna method, the Doppler method measures the
Doppler shift of the scattered radar echoes at oblique incidence (Figure 6).
The complex autocorrelation or the spectral analysis yields radial velocities
V' at different pointing directions. Assuming that the antenna beam points at
a zenith angle δ and at an azimuth angle α, directed clockwise from north, then

$$V' = U \cdot \sin\alpha \cdot \sin\delta + V \cdot \cos\alpha \cdot \sin\delta + W \cdot \cos\delta. \tag{8}$$

For a constant zenith angle and if the wind field is uniform, V' will vary
sinusoidally with the azimuth angle α. The Fourier analysis of the data series
$V'(\alpha)$ directly yields the eastward wind U and the northward wind V from the
Fourier coefficients, as well as the vertical velocity W from the constant off-
set. This is the so-called VAD method (velocity-azimuth-display).

This method can be simplified by using only three fixed beam positions,
one at $\delta = 0°$ and the other at a zenith angle $\delta \neq 0°$ and two azimuth angles α,
which preferably should be orthogonal to each other. We directly can solve (8)
to obtain U, V and W after measuring V' at the three beam directions. A most
simplified method uses only two beam directions at different azimuths and as-
sumes that the vertical velocity is negligible. We will call these the fixed-
beam methods. Two or three fixed-beam positions are mostly used in MST radar
applications. With fixed beams, however, most preferably the vertical, and

off-vertical directions in four orthogonal azimuth directions at equal zenith angles should be used. This would allow to reduce errors due to non-uniformity and inclined structures as well as to measure uniquely the correlation $\overline{U'W'}$ of the horizontal and vertical velocities (see section 6.3).

To obtain a most accurate estimate of the vertical velocity W, a velocity elevation display (VED) can be used. Here the azimuth is kept constant and the elevation is changed from $+\delta$ to $-\delta$. A fit of the portion of the sinusoidal variation within these limits of δ to the data series $V'(\delta)$ yields $U \cdot \sin\alpha + V \cdot \cos\alpha$ and W. Extending this elevation scan to two azimuth directions, also U and V can be determined.

Further information on the VAD, VED and fixed-beam Doppler measurements can be found in WILSON and MILLER (1972), BATTAN (1973), KOSCIELNY and DOVIAK (1983), STRAUCH et al. (1983).

An example of a VED measurement with an MST radar is shown in the series of spectra shown in Figure 8 (from ROTTGER et al., 1981). The antenna beam (half power beam width = 1.7°) was swung in the zonal plane ($\alpha = 90°$) in zenith angle steps $\Delta\delta = 3.4°$. The Doppler shift increases proportional to sin δ, which is obviously expected if the zonal wind component U is non-zero. The remaining Doppler shift at $\delta = 0°$ (e.g. at 17.2 km) is due to a vertical velocity W. Evidently, the signal power decreases with δ, which indicates that the detected anisotropic structures are horizontally stratified. Within these limits of δ the signal power decreases about exponentially by 1-2 dB/degree in the troposphere and 2-3 dB/degree in the lower stratosphere. This effect is also called aspect sensitivity.

The spectral width $\overline{\omega}_w$ is a measure of the turbulent velocity fluctuations in the radar volume. However it is diluted by the effects of beam width and wind shear broadening (e.g. BATTAN, 1973; HOCKING, 1983a). The former occurs because the radial components of the wind velocity vary from one edge of the radar beam to the other. The effect of beam width broadening is the stronger the wider the beam width is. The latter occurs if a shear of the wind
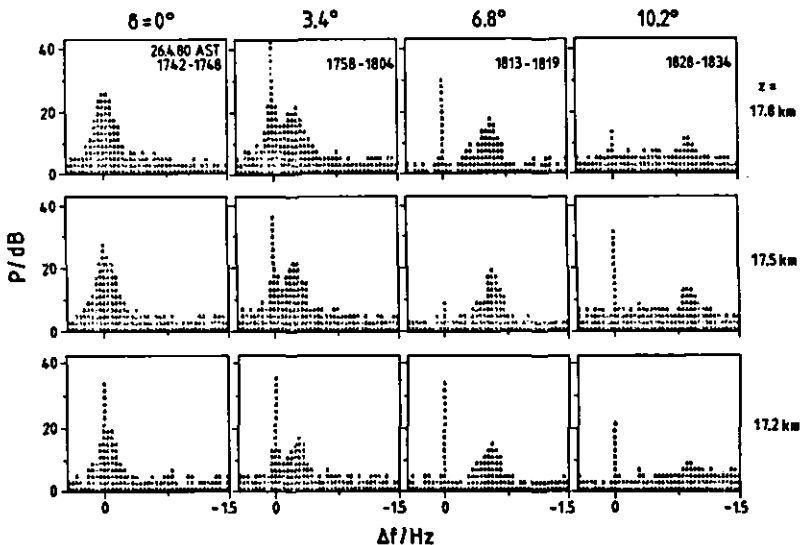

Figure 8. Spectra measured with an elevation scan fixed azimuth.

exists across the width of the beam or along the radial extent of the radar
volume (because of the finite width of the range gate). Vertical wind shears
or gradients additionally create an offset of the mean velocity if the extent
of the range gate Δr is larger than the gradient scale length (SATO and FUKAO,
1982).

The finite beam width also gives rise to errors in estimating the Doppler
shift since the scattering/reflection process is not isotropic. The product
of the antenna beam pattern with the angular distribution of the aspect sensi-
tivity yields an effective beam direction which is closer to the zenith than the
real pointing direction. This error is largest if a strong reflected component
is observed and can result in a velocity underestimate of some ten percent. It
can be minimized by applying the measured aspect sensitivity for correction.
Even for scattering, a similar underestimate can arise if there are strong hori-
zontal fluctuations of the wind velocity. Here the width of the spectrum de-
pends on the zenith angle, and it can be shown (HOCKING, 1983a,b) that the
contributions from subvolumes at different zenith angles within the antenna
beam result in a bias towards low frequency components of the spectrum, i.e.
yielding an underestimated horizontal velocity. Depending on the direction and
the suppression of antenna sidelobes, the spectrum will also be altered which
results in errors of power, mean velocity and velocity fluctuations.

As a result of these considerations, the width of the antenna beam and of
the range gate should be made as narrow as possible. Of course, the sampling
rate has to be made short enough to resolve the full shift and width of the
signal spectrum.

There are several reasons to choose the zenith angle δ of the antenna beam
as small as possible: 1) The spatial uniformity of the wind field is better
for smaller separations of the probed radar volumes. Deviations from uniformi-
ty cause a higher bias and variance of the velocity estimates, especially if
only fixed beam methods are used (KOSCIELNY and DOVIAK, 1983). 2) The altitude
resolution Δz diminishes with increasing zenith angle δ. This is fairly pro-
nounced for broad antenna beam widths. 3) Since the range r to a specified
altitude increases with zenith angle, a loss of sensitivity results due to the
r^{-2} dependence of the echo power. 4) The effective aperture decreases with
zenith pointing angle for a fixed antenna aperture assembled in the horizontal
plane. 5) The anisotropy of the scattering/reflecting process yields stronger
echoes at near-zenith angles.

Other effects, however, support the choice of large zenith angles: 6) The
uncertainty of the velocity estimate $\Delta V'$ decreases with increasing Doppler
shift, because the ratio $\overline{\omega}_D/\overline{\omega}_w$ decreases with δ (assuming that $\overline{\omega}_w$ varies
less with δ than $\overline{\omega}_D$) and $\overline{\omega}_w$ is consistent with the variance of $\overline{\omega}_D$. 7) For
a given $\Delta V'$, the uncertainty of the horizontal velocity is $\Delta V = \Delta V'/\sin \delta$.
However, the larger the zenith angle the shorter the preintegration time must be
chosen to avoid aliasing and comb-filter attenuation. 8) An instrumental error
in the pointing direction causes larger deviations of the horizontal velocity
estimate for small δ. 9) The effective pointing direction due to the aspect
sensitivity is assumed to become negligible at large zenith angles where iso-
tropic scattering will dominate.

In summarizing these arguments we find that a general proposal cannot be
made to select an optimum pointing angle. As in the spaced antenna method
(where one has to preselect antenna spacings) the beam pointing angles have to
be preselected when applying the Doppler method. The application of either the
spaced antenna and/or the Doppler method has advantages and limitations, and
one has obviously to deal with a trade-off between accuracy, simplicity and
cost-effectiveness when planning and operating an MST radar system.

Comparisons of Doppler wind measurements with rawinsondes were done with almost all existing radars. A fair-to-good agreement was always reported (e.g., LARSEN, 1983c). As an example we show in Figure 9 comparisons of radar winds measured with the UHF radar at Arecibo and rawinsonde winds measured at San Juan. FUKAO et al. (1982) state that most of the differences in the lower stratosphere can be explained by experimental errors, especially those of the rawinsonde. The difference in the upper troposphere wind data are explained by the spatial and temporal variations in the wind field. FUKAO et al. come to the agreeable conclusion that UHF and VHF Doppler radar measurements of winds provide a greater frequency and accuracy than the use of conventional rawinsondes. We also have to add that these radars can measure the vertical velocity with better accuracy than the yet applied methods, which is of considerable interest to meteorologists.

6.3 Radar Interferometry

Both, the Doppler and the spaced antenna drift method, do not separately evaluate the spatial distribution of the phases of the field pattern at the ground. With the spaced antenna set-up the amplitudes and the phases can be measured. Combining in a suitable procedure the complex signals from different antennas is in a wide sense the application of the interferometer technique. In the first spaced antenna measurements with VHF radars, ROTTGER and VINCENT (1978) and VINCENT and ROTTGER (1980) applied this method to measure the angular spectrum of tropospheric returns. It can be expected that further valuable additional information on structure shapes and motions as well as on atmospheric waves can be obtained with this technique.

An essential point which has to be stressed at first is the improvement of the vertical velocity measurements by using a spaced antenna interferometer. The basic principle of the technique is sketched in Figure 10. Let us assume diffuse reflection from a rough surface or structure S which is sufficiently far from the radar antenna and which is slightly tilted to the horizontal by an
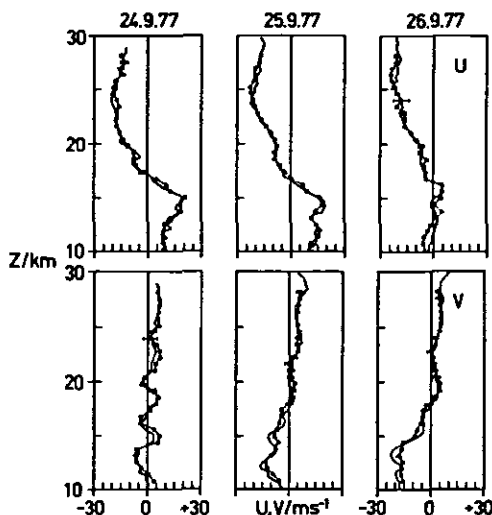

Figure 9. Mean zonal (U) and meridional (V) wind velocities measured with the Arecibo radar (dots) and rawinsondes (after FUKAO et al., 1982).
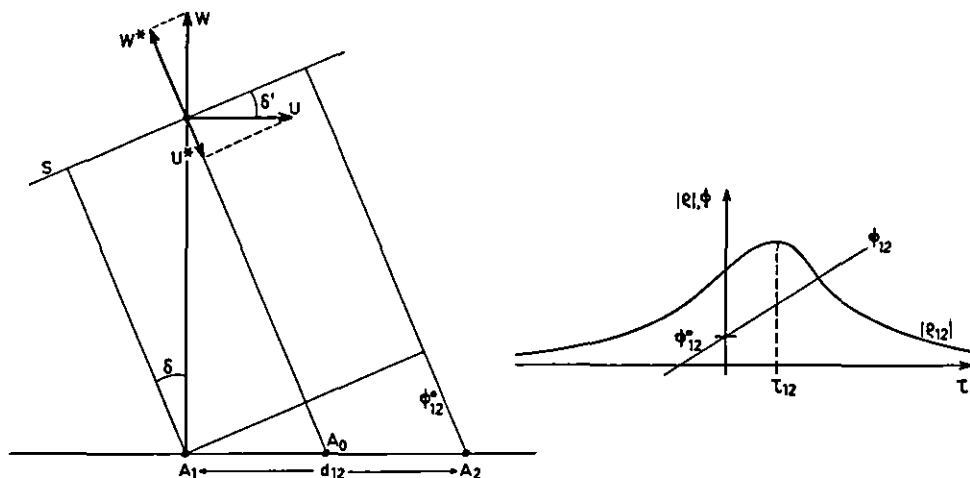
Figure 10. The principle of phase measurements
at antennas A_1 and A_2.

angle δ'. This structure moves with a velocity given by the horizontal component U and the vertical component W. A radar with vertically pointing antenna A_o with beam width larger than δ' measures the radial velocity

$$V' = W^* - U^* = W \cdot \cos \delta' - U \cdot \sin \delta'.$$

Thus, even when knowing the horizontal velocity U, the vertical velocity is still incorrect, if δ' is unknown (ROTTGER, 1981c).

The reflected signal can also be received at two separate antennas A_1 and A_2, and the complex cross correlation function ρ_{12} be computed. Its amplitude $|\rho_{12}|$ and phase ϕ_{12} is sketched in the lower part of Figure 10. From the displacement τ_{12} of the maximum of $|\rho_{12}|$ and the horizontal separation d_{12} of the receiving antennas, the apparent velocity $V_a = d_{12}/2\tau_{12}$ is calculated. We assume here for simplification that $V_a = U$ (instead of V_a, the true velocity has to be calculated according to the considerations in section 6.1 and BRIGGS in this Handbook). The radial velocity is calculated from the time derivative of ϕ at $\tau = 0$:

$$V' = \lambda_o/4\pi \cdot d\phi (o)/dt.$$

The tilt angle δ', which is similar to the incidence angle δ, is

$$\delta' = \arcsin(\phi_{12}(0) \cdot \lambda_o/2\pi d_{12}).$$

This yields the corrected vertical velocity

$$W = (V' + V_a \sin\delta)/\cos\delta.$$

For a typical ratio $U/W = 100$ and $\delta = 0.6°$, for example, the vertical velocity estimate would be incorrect by a factor of 2 (0.5) if this correction were not applied. The angle δ' is equivalent to the inclination of the reflecting structures. Its average can give an estimate of the inclination of isentropic

surfaces (baroclinicity), which is evidently of interest for meteorological applications (GAGE, 1983).

A phase lag between two antennas can also be introduced instrumentally, either by including a delay line between the antennas and the receiver or by adding a phase lag to the complex signal samples received at the different antenna channels. In the first case (preselected beam direction) the signals from both antennas would be added in an analogue coupler, and in the second case (post-selected beam direction) the digital samples would simply be added as vectors during data processing. This procedure is equivalent to electrically swinging the beam to different angles δ. The addition (distribution) of signals from (to) different antenna modules is used to swing the receiver (transmitter) beam of a phased array. For separate transmission and receiving antennas, the transmitter antenna beam can be kept fixed and the receiving antenna beam be swung by post-selection as long as the angle δ is narrower than the beam of the transmitter antenna. The advantage of this latter method is the post-selection of all possible angles δ, whereas the former method preselects an angle which cannot be changed after the data were taken. Of course, this method is easily extendable from the explained 1-dimensional example to a 2-dimensional application as well as to an array of more than two receiver antenna modules.

This radar interferometer method allows not only the measurement of the horizontal wind components by the Doppler method, but also the tilt, the aspect sensitivity and the horizontal phase velocity and wavelength of atmospheric waves.

VINCENT and REID (1983) applied two symmetrical fixed beam directions at zenith angles δ and −δ with an HF radar to study mesosphere dynamics. They showed that this generally allows measurement of the cross correlation $\overline{U'W'}$ of the fluctuations W' and U' of the vertical and horizontal velocities, respectively:

$$\overline{U'W'} = (\overline{V'^2}(\delta) - \overline{V'^2}(-\delta))/2 \sin 2\delta.$$

It is evident that this comprises a very powerful tool to measure gravity wave parameters and momentum flux, which is proportional to $\overline{U'W'}$.

Originally, Vincent and Reid used two fixed transmitter/receiver antenna beams. It is also possible to apply the method of post-selecting the receiver beam directions. This is illustrated by one example shown in Figure 11. Here the beam of the spaced receiving antennas of the SOUSY-VHF-Radar was "steered" to δ = +1.2° and δ = −1.2° by digitally inserting a phase delay to the recorded data. This allowed observation of two different phase locations of a wave in the stratosphere, which becomes clear due to the phase shift between the two time series of the displayed radial velocity. From these measurements, either by using post- or preselected beam directions, the horizontal wavelength and phase velocity, the momentum flux and the mean horizontal and vertical velocities can be deduced. The full 3-dimensional information has to be obtained by steering the antenna in two orthogonal vertical planes. The vertical wavelength and phase velocity can be found from just one, preferably vertical beam direction. There are also influences on spectral width and received power due to gravity waves, which were analyzed by GAGE et al. (1981).

Another application, which originally was used to study ionospheric plasma turbulence (FARLEY et al., 1981), was recently also applied to MST radars to trace discrete structures, such as blobs of turbulence moving through the antenna beam (ROYRVIK, 1983b; IERKIC and ROTTGER, 1984). By making use of a cross spectrum analysis and observing the change of δ and V' as a function of time, not only the location of the blobs but also their vertical and horizontal velocities can be measured more accurately.
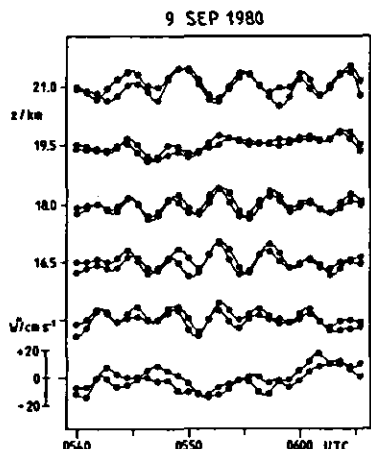
9 SEP 1980

Figure 11. Vertical velocity oscillations due to a gravity wave, measured at two slightly different zenith directions with the interferometer technique.

The measurements of δ and its temporal variations are very useful to avoid erroneous interpretations that the radial velocities measured with vertical antenna beams are really vertical velocities. All these techniques, evaluating the temporal as well as the spatial variations of amplitude and phase of the field pattern, we may call MST radar interferometry. Their applicability to MST radars has already been tested, but more refinement is needed to fully exploit this promising technique.

7. MST RADAR ANTENNAS

7.1 Basic Parameters

The considerations on the interferometer applications can be extended into a brief description of antenna arrays (e.g., SKOLNIK, 1970). Assume that N individual antenna elements with equal spacing d are horizontally lined up to form a multi-element array. In the array far-field ($r > (N \cdot d)^2/\lambda_o$) the polar diagram is

$$E(\delta) = \sum_{n=1}^{N} E_n(\delta)\exp(i(-\frac{2\pi(n-1)d}{\lambda_o} \sin\delta + \phi_n)), \tag{9}$$

where $E_n(\delta)$ is the pattern of an individual element and ϕ_n is a relative phase placed on this element. Thus, the polar diagram is just the Fourier transform of the spatial aperture distribution. If all the elements have similar $E_n(\delta)$ and ϕ_n, the polar diagram is a function sin Nx/sin x with $x = \pi d \sin\delta/\lambda_o$. If $d < \lambda_o$, there is only one main lobe at $\delta = 0°$ (for $\phi_n = 0$). If $d > \lambda_o$, grating lobes at $\delta_G = \arcsin(\lambda_o/d)$ occur, for $E_n(\delta) = $ const with amplitudes similar to the main lobe. The width of the main lobe is $\delta_B = \arcsin(\lambda_o/Nd)$. In radar applications also the two-way beam width $\delta_B/\sqrt{2}$ is used. If δ_B is small, it is directly proportional to the ratio of the wavelength λ_o to the aperture dimension N · d. Nulls of the radiation pattern are at $\delta = \pm n \cdot \delta_B$, where n = 1, 2, ..., N/2, and sidelobes occur at $\delta_s = \delta_n \pm \delta_B/2$. If equal weighting W_n is applied to each of the single elements (e.g., $E_n(\delta) = W_n \cdot E_n'(\delta)$, with $W_n = $ const for all n), the first sidelobe closest to the main lobe is suppressed by 13.2 dB plus the fall-off of $E_n'(\delta)$ at δ_{s1}.

The antenna pattern, namely the direction δ_o of the main lobe, can be changed by applying a linearly progressing phase ϕ_n from element to element, which has to be

$$\phi_n = 2\pi d(n-1)\sin \delta_o/\lambda_o.$$

This beam steering should be within reasonable limits of the individual element pattern $E_n(\delta)$ to avoid undesirable degrading of the antenna radiation.

To obtain improved sidelobe suppressions, a tapering of the antenna array can be applied by either changing the weighting function W_n (electrical weighting, i.e. feeding the outer elements of an array with less power than the inner elements), or by using unequal element spacings d_n (spatial weighting, i.e. applying larger spacings for the outer elements). The price one has to pay for the improvement of sidelobe suppression by tapering is always a broadening of the main lobe and a lowering of the antenna gain G. Using a triangular weighting, for instance, improves the one-way, first sidelobe suppression to -26 dB, but widens the main lobe by a factor of 1.44 and reduces the gain by 25%, as compared to uniform weighting. The respective values for a $\cos^2$-weighting, which is a good approximation to a Gaussian weighting, are -32 dB, 1.64 and 33%.

These 1-dimensional considerations can be extended easily to a realistic 2-dimensional radar antenna array by using, instead of d, the projections d' of all element positions onto an axis elongated in the azimuth direction α. The total antenna pattern then can be calculated by (9) for any α and δ. For a real radar antenna we also have to consider that the radiation can be into only one half-sphere. For an array system this means that reflections from the ground, a screen or reflector elements have to be included in the calculations.

The antenna gain G is defined as the ratio of the maximum radiation intensity (in the main beam) to the average radiation intensity (averaged over all δ and α). For an antenna array with reasonable sidelobe suppression it is given by the antenna area or approximately by the inverse of the two orthogonal beam widths δ_B' and δ_B'' (in radians):

$$G = \frac{4\pi A}{\lambda_o^2} \sim \frac{4\pi}{\delta_B' \, \delta_B''} \quad . \tag{10}$$

The effective area A or aperture of the antenna is the product of the physical area of the antenna and the efficiency of its illumination, which for instance may be reduced by tapering. It is noted that A does not include the losses of the antenna and its feed system.

The considerations of antenna arrays, consisting of several discrete elements, can generally be extended also to antennas with continuous aperture illumination, such as dish antennas. The aperture A of a phased array or a dish antenna is used to calculate the power-aperture product $P \cdot A$, which defines the sensitivity of an MST radar. Note that P is the effective average power, which is radiated by the antenna, i.e. it is smaller than the output power of the transmitter due to losses in the antenna and feed lines.

This very brief outline is useful to estimate some basic values needed to plan and design a VHF-radar antenna system. The minimum requirement for investigations of the troposphere and stratosphere is a power aperture product $P \cdot A = 10^6$ Wm^2 (see Figure 15) where $P = P_a$ is the average transmitter power. We assume that the radar sensitivity is a function of A^k, where $1 < k < 2$ depending on the reflected contribution (see equation (1)). It means that in practice one should favor an extension of the antenna area against an increase of the transmitter power. For a commonly achievable transmitter power $P = 10^3$W, $P \cdot A = 10^6$ Wm^2 yields the dimension of a circular antenna array to be about $N \cdot d \approx 36$ m, and $A \approx 1000$ m^2. For $\lambda_o = 6$ m we can obtain a half-power beamwidth $\delta_B \approx 9°$, and a gain $G \approx 27$ dB. This estimate assumes that the array is optimally filled with elements. A single dipole element over a proper reflector screen has an effective area of about 15 m^2 (at $\lambda_o = 6$ m). Thus, about 64 dipoles are needed to fill the array. For a square array with 8 x 8 elements the resulting spacing is d $\approx 0.75 \lambda_o$, and a grating lobe will not

occur. The first nulls are at $\delta_n = \pm 9°$, and the first sidelobes at
$\delta_s = \pm 13.5°$. To use such an array with the Doppler method, the antenna beam
has at least to be steered to a zenith angle $\delta_c = 9°$ to place a null into
zenith direction (minimizes influences of aspect sensitivity). Even then a
sidelobe at $-4.5°$ is a problem.

7.2 Sidelobe Effects

When designing an MST radar antenna one has to trade between the choices to
optimize the effective aperture or to optimize the sidelobe suppressions. An
optimization of the aperture increases the sensitivity. Suppression of side-
lobes by tapering, attenuates undesirable signals which spoil the estimates of
reflectivity and velocity. Generally, any sidelobe effects are equivalent to a
broadening of the antenna beam. The return signal is due to a product of the
antenna pattern with the varying atmospheric reflectivity structures. Thus,
knowing the antenna pattern, it is in principle possible to find the signal
spectra, which however may be a tedious computational and ambiguous procedure.

For vertically pointing main beams the sidelobe effects are efficiently
suppressed because of the aspect sensitivity. It follows that sidelobes are a
minor problem for spaced antenna methods. However, they can be crucial for
Doppler methods, which need off-vertical beams. If a sidelobe is pointing to-
wards the zenith a larger power may be received from the vertical than from
off-vertical directions but quantitative estimates of this effect are not yet
known.

To get an error estimate of sidelobe effects with an off-vertical main
beam, we discuss the following 1-dimensional example. This yields a reasonable
estimate since the sidelobe closest to zenith, in the plane in which the main
beam is steered, mostly dominates the errors. For simplification we also as-
sume that the width of the main beam and sidelobes are much narrower than the
chosen off-zenith pointing angles. The general outcome of the error estimates,
however, will not be altered considerably for wider antenna lobes. Let P_1 be
the power gain of the main lobe at δ_1, and P_2 the power of a sidelobe at δ_2.
The aspect sensitivity shall be given by $a(\delta)$, and the radar detected structures
shall move with a horizontal wind $-U$. Then equations (6) or (7) yield:

$$\overline{P} = a_1 P_1 + a_2 P_2,$$

$$\overline{\omega}_D = \omega_1 a_1 P_1 / (a_1 P_1 + a_2 P_2) + \overline{\omega}_2 a_2 P_2 / (a_1 P_1 + a_2 P_2).$$

With $a_{1,2} = a(\delta_{1,2})$ and $\overline{\omega}_{1,2} = 4\pi/\lambda_o \, U \sin \delta_{1,2}$, we obtain the frac-
tional errors $\Delta \overline{P}$ and ΔU, if one would neglect the sidelobe effects:

$$\Delta \overline{P} = (\overline{P} - a_1 P_1)/a_1 P_1 = a_2 P_2 / a_1 P_1$$

and
$$\Delta U = (\overline{\omega}_D - \overline{\omega}_1)/\overline{\omega}_1 = (a_1 P_1 + a_2 P_2 \sin\delta_2/\sin\delta_1)/(a_1 P_1 + a_2 P_2) - 1.$$

Using $\delta_1 = 9°$, $\delta_2 = -4.5°$ (without considering the apparent beam direc-
tion), $P_1 = 1.0$, $P_2 = 10^{-2}$ (for a worst estimate of -20 dB two-way beam
sidelobe suppression) and $a(\delta) = 1$ dB/degree corresponding to $a_1 = 1.3 \cdot 10^{-1}$
and $a_2 = 3.6 \cdot 10^{-1}$, yields $\Delta \overline{P} = 2.8 \cdot 10^{-2}$ and $\Delta U = -4.3 \cdot 10^{-2}$. This
means that the power would be overestimated by 2.8%, and the horizontal velocity
would be underestimated by 4.3%. A similar computation will yield the error
estimates for the spectral width. Since δ and $P_{1,2}$ are known instrumental
parameters and $a(\delta)$ is fairly well known from observations, $\Delta \overline{P}$ and ΔU can be
used for correction, which will yield a very reasonable estimate of $\overline{P}$ and U.
It, thus, may appear to be more feasible to apply corrections, which have to
take into account the two-dimensional pattern, than use tapering. However,
situations may occur (e.g., strongly tilted, reflecting layers), which would
lead to substantial contributions through antenna sidelobes. These situations
need special attention during antenna design and data analysis.

Also sidelobes at low elevation angles have to be considered since these cause strong echoes from the non-atmospheric targets in the surroundings of the radar antenna. It can be shown that the ground reflection very effectively cancels the direct radiation at grazing angles ($\delta \gtrsim 85-90°$). because the reflected wave suffers a phase reversal during reflection. This even can suppress low sidelobes of the array pattern which may be regarded as crucial without taking into account ground reflections. The location of an array antenna at a flat ground (extending out to several 100 m) may be sufficient but a shallow valley generally should be preferred to further suppressing the low angle radiation effects. However, high extending targets, such as radio towers or mountains in the close vicinity, will still cause considerable clutter echoes, even when optimizing the antenna array for low angle radiation suppression.

7.3 Antenna Types and Feed Systems

VHF radars generally operate with quasi-vertical beams. i.e. the zenith angles are smaller than about $20°-30°$. For MST work linear polarization is sufficient. Essentially four different types of antenna systems are in use: dish antennas. dipole arrays. coaxial-collinear arrays (coco) and Yagi arrays (e.g. Figure 12).

Dish or cylinder antennas are rarely used for VHF radar applications be- cause of their large dimensions (NAIC, 1976; ROTTGER et al., 1981; HAGFORS et al., 1982). Only one or a few elements are applied as primary feed antennas. The beam steering is done either by moving the position of the feed antenna or the entire dish. This has the advantage that no complicated power distribu- tion and phasing network has to be used to feed the antenna. Because of the
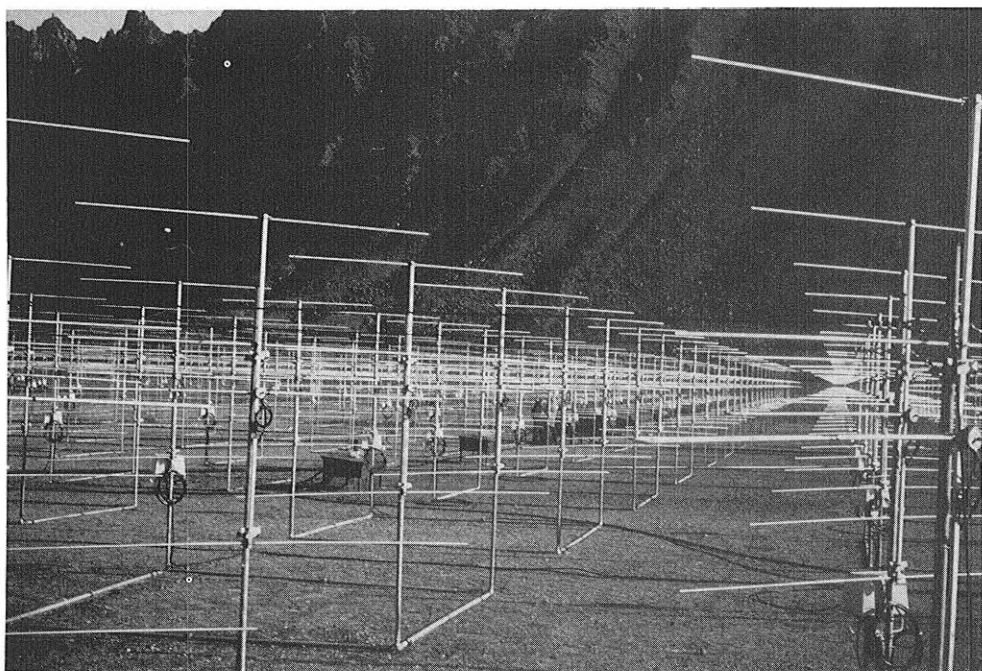

Figure 12. Yagi antennas of the SOUSY radar on Andoya/Norway (photo: P. Czechowsky).

limited size of the primary feed antennas. the low-angle sidelobe suppression
is not sufficient, which results in strong clutter echoes.

All other phased array antennas consist of many single elements which have
to be fed by a cascading network of cables. hybrids and phase shifters. The
coco antenna (OCHS. 1965; BALSLEY and ECKLUND. 1972) has the advantage that the
feeding of elements in one line is just done by interchanging the inner and
outer conductor of a coaxial cable every half wavelength. The outer conductors
of the aligned cable act as collinear dipoles. The feeding is done from the
center of a line. which may typically consist of 16-48 dipoles. Positioning
several of these strings or rows in parallel at spacing $d < \lambda_o$, and feeding
these by a suitable matching network results in a coco array. The radiation
and the losses in a coaxial string comprise a natural tapering, suppressing
sidelobes in the plane of the string but also degrades the radiation efficiency.
Because of the phase relation along a string is fixed. a beam steering parallel
to the coco string is not possible. Beam steering is achieved by inserting
appropriate phase delays in the cables feeding the parallel rows. The coco
dipoles have to be lined up a quarter wavelength over reflector wires or
screens. This antenna type is cheap, because coaxial cable is used for
antenna elements and the matching network is simple. The successive phasing
from one collinear element to the next. however, degrades the bandwidth of
this antenna type. Instead of coaxial cable as radiator, half-wave dipoles can
also be used which are fed in a properly adjusted phase to form a collinear
antenna. The application of collinear dipole lines limits the steerability of
an array, and for this reason frequently three antenna arrays are used with
three different fixed beam directions (see GREEN et al., 1975; BALSLEY et al.,
1980).

Single dipole or Yagi antenna elements. forming an array, are fed by a
cascading network of open wire or coaxial cable systems (e.g., CZECHOWSKY and
MEYER. 1980; CZECHOWSKY et al., 1984). The cascading is most appropriately
done in 2^n branchings (n = 1.2,...). By these means one can feed parts of
the antenna array with 1/2. 1/4,... power to provide tapering without dissipa-
ting power. The branching is best done in couplers. power dividers or hybrids,
which prevent power, reflected from a mismatch to return to the transmitter, or
other antenna elements, and also minimize effects due to mutual coupling between
the single antenna elements. The coupling can be critical if phase control is
applied to steer the antenna beam. The phase control is usually inserted close
to the final elements, which may be connected to form modules of 4 or 16
elements. For continuous beam steering, phasing is most easily done by
inserting discrete phase delays in steps of $2\pi/16$ with a binary phase shifter.
This can be achieved by only four relays, switching phases of 22.5°, 45°. 90°
and 180°. The advantage of Yagi against dipole antenna elements is that no
ground screen is needed because of the Yagi reflector. The multi-element
structure of a single Yagi allows for a higher gain (improving the filling
factor of an array) and a negligible coupling (<-25 dB) between adjacent Yagi
antennas in an array. Mostly Yagi antennas can be constructed in such a way
that the bandwidth is several Megahertz. The bandwidth limiting factors in a
Yagi system essentially are the phase shifters. The losses of a Yagi system are
also considerably lower than those of a coco antenna. However, such a Yagi
system is obviously more expensive than a coco system. The feeding of an
array system can be from one transmitter. but also sub-modules can be fed
separately by several phase-controlled transmitters (FUKAO et al., 1980). The
transmitter phase control can even be used to steer the antenna beam. This,
however, needs a similar phase control of the receiver channels.

8. MST RADAR EQUIPMENT

8.1 Transmit-Receive System

Separate antenna arrays can be used for the transmission and for the reception mode, but it is more effective to use only one antenna for both modes. While separate antenna arrays allow for sufficient decoupling of the receiver from the transmitter, a fast and highly insulating transmit-receive duplexer has to switch the antenna from the transmitter to the receiver and vice versa, if only one antenna is used. The response and recovery time of the duplexer should be in the order of the range sampling time, i.e. typically less than 10 μs. To insulate a peak transmitter pulse amplitude of 10^5W to a fraction of Watt, which will not destroy the receiver, the decoupling attenuation has to be better than 60 dB. These specified values can be obtained with pin-diode hybrid switches. For more details, the reader is referred to the technical report by CZECHOWSKY and MEYER (1980), where also more details on power distribution, phasing and antenna diagram calculations can be found.

A block diagram of a typical transmitter/receiver system of a VHF radar is shown in Figure 13. It depicts the design of the portable SOUSY Mini-VHF-Radar which·was operated at the Arecibo Observatory (ROTTGER et al., 1981). According to individual requirements, many variations and modifiations of such a system can be done. but Figures 13 and 14 describe a generally adopted design. based on the outline of Figure 3. The 120-MHz signal of the master oscillator (MO) is divided by four to obtain a 30 MHz intermediate frequency which can be modulated (MD) in amplitude (on/off) and phase (0°/180°). This is to provide phase coding and dc-elimination. A similar divider generates the 0° and 90° signals, which are necessary for quadrature detection. The operational frequency 46.8 MHz is generated by mixing with the local oscillator (LO) signal at 76.8 MHz. It is amplified in the transmitter (TX) and fed through the transmit-receive duplexer (TRX) and a reflectometer to the antenna. The signal received at 46.8 MHz is amplified in the receiver (RX), mixed with the LO-signal to an intermediate frequency (IF) of 30 MHz, and mixed down to the baseband in the quadrature detector (X). The two quadrature outputs, the real part (Re) and the imaginary part (Im), are low pass filtered (LP) to match the receiver bandwidth to the bandwidth of the transmitted pulse.

There are some basic system specifications: The stability of the oscillators should be better than 10^{-4} Hz/ms to allow a good accuracy of coherent detection. The transmitter peak power typically is between 1 kW and 1 MW with duty cycles up to several percent. The transmitter bandwidth must cover the shortest pulse length of 1 μs ($\hat{=}$2 MHz bandwidth). The transmitter is normally operated in class-C mode to gain an optimum efficiency.

The receiver noise figure need not be better than a few dB (some 100 K), since the sky noise level (> 1000 K) determines the sensitivity. A noise calibration signal (CAL) of 1000 K. say, should be injected into the receiver front-
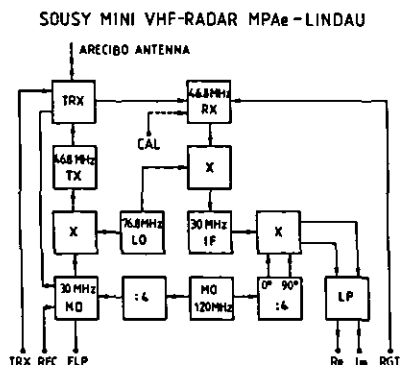
SOUSY MINI VHF-RADAR MPAe-LINDAU

Figure 13. Example of a transmitter-receiver system (after ROTTGER et al., 1981).

end in order to provide an absolute power calibration (which also needs a continuous monitoring of the transmitter power). The receiver linearity range shall exceed 60 dB in order to avoid saturation with strong clutter signals. The phase flip and the quadrature detection must be within less than a few degrees accuracy, and the amplitude ratio of the quadrature components must not deviate from unity by a few percent. Of course, the receiver base-bandwidth (postdetection) has to be about 1 MHz to allow detection of the shortest pulses. Bessel filters are usually used as low-pass or post-detection filters.

8.2 Radar Control and Data Acquisition

The control pulses for the transmitter/receiver and the ADC are generated in the digital part of the system where also the data acquisition takes place (Figure 14). In the displayed example no host computer is used. The system is simply controlled by preselected instructions from programmable read-only memories (PROMs) and by external settings. The dashed lines enclose those subunits of instructions, commands and data transfer which also could be executed *more flexibly by a host computer.* Typically the radar controller generates pulse trains of control pulses such as demonstrated in this example:

TRX

RGT

RFC

FLP

ADC

CAL   on / off
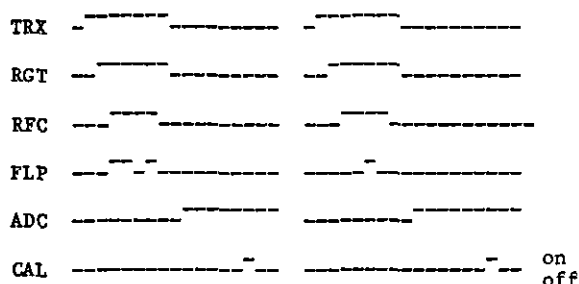
The transmit-receive duplexer (TRX) is switched on first, followed by a receiver gate pulse (RGT) and the switch-on of the transmitter radio frequency (RFC), which is phase flipped (FLP) between 0° to 180° for coding. The anallogue-digital-converter starts sampling (ADC) after RFC off and the TRX and RGT have opened the receiver. At the sixth sampled range gate a calibration signal (CAL) is injected. The whole sequence is repeated after one interpulse period when only the phase flip is inverted (for dc-elimination). For application of the complementary coding scheme a second double pulse code (consisting of the complementary pattern) is transmitted, and this radar cycle is repeated N times to yield one radar burst (section 4.2). Additionally, pulses for antenna con-
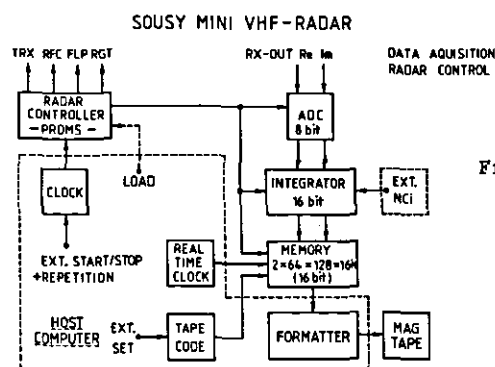
SOUSY MINI VHF-RADAR

Figure 14. Example of the digital part of an MST radar system. The functions of the parts in the dashed boxes are mostly executed by a computer (after ROTTGER et al., 1981).

trol and other purposes can be generated, changing from one to the next radar
burst. In more advanced systems multichannel receivers and ADCs are used (e.g.
spaced antenna parallel processing, e.g. ROTTGER, 1981a). It is generally
found that 8-bit ADCs are sufficiently matched to the linearity range of the
receivers (> 60 dB). If strong clutter signals are present the capacity must
eventually be extended to 12 bit or a range-dependent attenuation has to be
used. Applying an 8-bit ADC and a 16-bit integrator allows to add samples from
at least $N = 256$ radar cycles in one channel (see section 4.3). The pre-
integration time then would be $t_i = 2N \cdot T_{IPP} = 128$ ms for $T_{IPP} = 250$ μs,
corresponding to a maximum unambiguous range of 37.5 km and a maximum resolvable
radial velocity of 12 ms^{-1} ($f_{Dmax} = 4$ Hz). Applying a 4-bit complementary
code and a 300-m range resolution would result in a transmitter duty cycle of
about 3%.

The radar operation, which is synchronized, started and stopped by ex-
ternal clock control, takes place in several nested sequences (section 4):
(1) the radar cycle, i.e. the transmission of one code unit of radar pulses with
preselected duration and the sampling of the real and imaginary signal at pre-
selected range gates (e.g. 128). (2) the integration cycle (burst), i.e. the
repetition and coherent integration of an externally selected number of radar
cycles. (3) record cycle, i.e. the repetition of a preselected number of inte-
gration cycles (e.g. 64) to form one total record which is stored in a 16-k
memory. After completion, the memory content (preintegrated raw data) as well
as other parameter information (tape code, time etc.) are dumped in one tape
record. The data analysis can be done off-line with a separate computer. This
versatile device of a VHF radar transmitter/receiver, radar control and data-
acquisition system is housed in four small cabinets. and with a suitable an-
tenna array forms a complete operational VHF radar. Additionally, a hardware
decoder. correlator or array processor can be attached (e.g., WOODMAN et al.,
1980). and the whole system can be most flexibly controlled and monitored by a
minicomputer. as done with most radars nowadays. This also would allow some
real-time raw data display and analysis as well as automated real-time transfer
of meteorological parameters via a telephone modem. These expansions evidently
depend on the individual necessities and requirements of the experimenters.
The described minisystem, however. is already sufficient for a fairly optimized
radar operation.

Descriptions of the early VHF radar systems were given by WOODMAN and
GUILLEN, 1974; GREEN et al., 1975; CZECHOWSKY et al., 1976; ROTTGER et al.,
1978; BALSLEY et al., 1980). and design considerations were summarized by
BALSLEY (1978a.b). GAGE and BALSLEY (1978) and BALSLEY and GAGE (1982).

9. SENSITIVITY REQUIREMENTS AND SOME OBSERVATIONAL EXAMPLES

9.1 Empirical Estimates

The sensitivity of a radar can be defined by the minimum detectable re-
flectivity. The detectability, which is the probability of discriminating a
signal from noise. is given by the ratio of the average signal power to the
noise power. multiplied by the square root of number of observations. The
average signal-to-nosie ratio (SNR) depends on the radar reflectivity and in-
strumental parameters and thus can be used to estimate the sensitivity of MST
radars for a given number of data samples.

The noise in the lower VHF band is essentially given by the sky noise.
Depending on antenna beam width and direction as well as on time. the sky noise
varies between about 1500 K and 8000 K at frequencies around 50 MHz. The noise
power P_n is defined as the average power of sky noise within a typical signal
bandwidth (here 3 Hz). The radar equation yields the signal power
$P \sim P \cdot A \cdot \Delta r \cdot C^2$, where $P = P_a$ (see section 2.3). The effective radar

reflectivity c^2 and hence the signal power and sensitivity cannot be determined easily from theory, as was pointed out in earlier sections. We, thus, will use an empirical approach to determine the sensitivity of VHF radars, which can supplement the detectability estimate deduced by GAGE and BALSLEY (1978).

In Figure 15 a typical tropospheric/stratospheric height profile of the signal-to-noise ratio SNR = $\overline{P}/P_n$ (in logarithmic units) is shown. Since SNR depends on the instrumental parameters P, A and Δr, the abscissa x has to be shifted accordingly. A useful way is to define the abscissa $x' = \log(P \cdot A \cdot \Delta r) - \log$ SNR, since this directly allows to determine $\log$ SNR $= \log(P \cdot A \cdot \Delta r) - x'$ for a given altitude z and given instrumental parameters P, A and Δr. These parameters are inserted in W, m^2 and units of 300 m, respectively. The solid curve $\tilde{S} = z(x')$ directly yields the altitude z_0 where the signal-to-noise ratio is unity; e.g., $z_0 = 20$ km for $\log(P \cdot A \cdot \Delta r) \approx 7$, which can correspond to $P = 10^4$ W, $A = 10^3$ m^2 and $\Delta r = 300$ m.
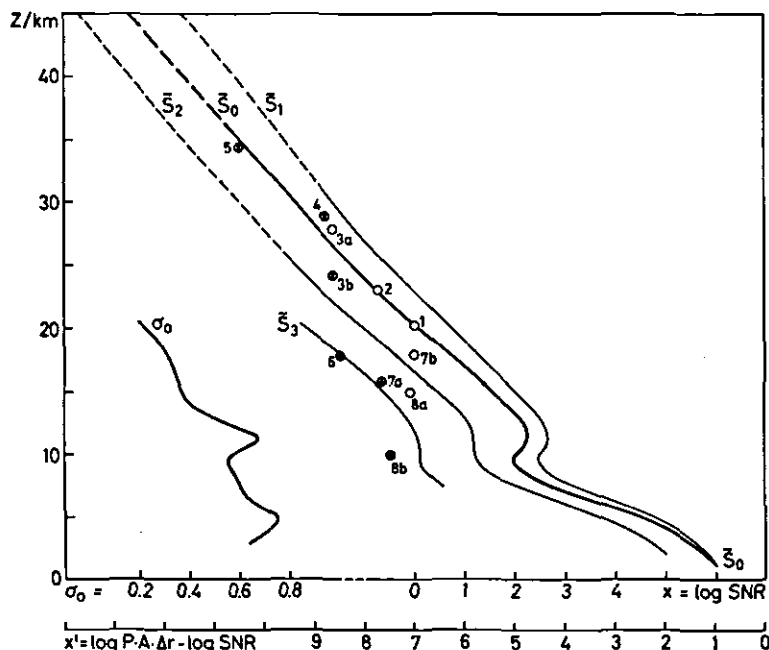

Figure 15. Average signal-to-noise as a function of altitude; the abscissa x is for $P \cdot A \cdot \Delta r = 10^7$. The abscissa x' gives log SNR for radars with given power-aperture-range gate product ($P \cdot A \cdot \Delta r$). $\tilde{S}_0$ is for $\delta_0 = 0°$. k=1.0 and A/P=0.1.
$\tilde{S}_1$: $\delta_0 = 0°$. k=1.25 ($z \gtrless 12$), k=1.2 ($z < 10$), A/P=1.
$\tilde{S}_2$: $\delta_0 = 7.5°$. $\bar{\Delta} = 0.15$/deg (z=12), $\bar{\Delta} = 0.1$/deg (z<10).
$\tilde{S}_3$: $\delta_0 = 15°$. The circles denote measured upper heights (log SNR=0) with different radars (1.2.3=SOUSY; 4=SOUSY-Arecibo; 5=Jicamarca; 6=Poker Flat; 7=Platteville; 8=Sunset). The open circles denote vertical antenna beams, the crossed circles Ø denote off vertical beams ($\delta_0 < 7°$), $\oplus$ denote $\delta_0 = 15°-30°$. σ_0 is the log normal standard dveiation of $\tilde{S}_0$.

The curve $\tilde{S}_o$ below 20 km was measured with the SOUSY-VHF-Radar in the spaced antenna mode (vertical beams). It represents an average over 290 hours of operation during 12 days between 28 Oct. and 12 Nov. 1981. This curve is characteristic for other VHF-radar observations in the middle latitudes. It shows the exponential lapse of received echo power with altitude and the typical increase of echo power above the tropopause. There are several regions with different echo lapse rate $\tilde{\Delta} = -\Delta \log SNR/\Delta z$, these are: the lower troposphere ($z < 4$ km) with $\tilde{\Delta} \approx -0.4$ km^{-1}, the middle and upper troposphere ($4 < z < 10$ km) with $\tilde{\Delta} \approx -0.5$ km^{-1}, the tropopause region ($10 < z < 12$ km) with $\tilde{\Delta} \approx +0.2$ km^{-1}, the lower stratosphere ($12 < z < 25$ km) with $\tilde{\Delta} \approx -0.25$ km^{-1} and the middle stratosphere ($z>25$ km) with $\tilde{\Delta} \approx -0.2$ km^{-1}. It cannot be excluded, however, that the middle stratosphere lapse rate is underestimated because of low signal-to-noise ratios of the measurements. The extrapolation of $\tilde{S}$ to larger altitudes than 20 km was done by using the lapse rates $\tilde{\Delta}$ from RUSTER et al. (1980) and BALSLEY and GAGE (1981).

Additionally, other limits are drawn in Figure 15. Since the signal power for reflection depends on A^k ($1 < k < 2$) instead of A, one expects for $P \cdot A = const$ a higher SNR if the ratio A/P increases. Thus, curves $\tilde{S}_o, \tilde{S}_1$ for two reasonable ratios $A/P = 0.1$ (e.g. SOUSY) and $A/P = 1$ (e.g. Jicamarca) are drawn. These two curves are valid only for quasi-vertical incidence. The aspect sensitivity $\overline{\Delta}$ reduces the SNR for off-vertical beams, which yields the lower limit curve $\tilde{S}_2$. A zenith angle $\delta_o = 7°$ and an angular dependence of log SNR of 0.1/degree in the troposphere and 0.15/degree in the stratosphere is used; $\tilde{S}_3$ is for $\delta_o = 15°$.

Data points from other measurements taken at different radars with different zenith angles are included. Some uncertainty remains on these data points, since the exact average power aperture product and/or the noise bandwidth was not available. The different measurements also suffer from the varying noise level which yields an uncertainty in log SNR of about 0.4 units. However, the measured points generally confirm that the sensitivity relation given by $\tilde{S}$ in Figure 15 can be used as an average standard.

The mesosphere echoes (CZECHOWSKY et al., 1979; FUKAO et al., 1979; ECKLUND and BALSLEY, 1981) are much more variable and cannot be predicted as easily; a brief description can be found in the following section and in Figures 16 and 19.

The effective reflectivity and hence the average sensitivity change due to variable strength of turbulence and refractive index structures, which depend on the meteorological conditions (e.g. NASTROM et al., 1982). Height variations of the tropopause, which may be as high as 18 km in the tropics, change the position of the secondary maximum around 10 km. The tropospheric reflectivity is higher when the humidity is increased. The average distribution of the logarithm of the reflectivity ($\hat{=} \log \bar{P}$) is in a good approximation log-normal with rms deviation σ_o. A height profile of σ_o is plotted in the left-hand corner of Figure 15, which was obtained from the 12-day observation period. It represents the general variability of the radar reflectivity, which decreases with height and is obviously higher in the troposphere (about ± 0.6 log-units) than in the stratosphere (about ± 0.3 log-units). The high σ_o values in the lower stratosphere are due to the fact that the tropopause height is fairly variable over a 12-day observation period.

9.2 Some Typical MST Radar Records

Figure 15 shows profiles which were obtained by averaging over many days, but obviously also a seasonal variation occurs. This is depicted in Figure 16, where we notice that the upper stratosphere height of detectable signals is

lower in summer than in winter. An explanation of this observation may be that the turbulence activity in the lower stratosphere differs with season (BALSLEY et al., 1983).

In Figure 17 an example of the medium-term variability (period of a day) of reflectivity is shown, which is characterized by some relevant structures. This contour plot, which is discussed in detail by ROTTGER and SCHMIDT (1981) and LARSEN and ROTTGER (1982), indicates the usefulness of VHF radar operations to observe the tropopause height and frontal boundaries. The reflectivity records combined with the observations of aspect sensitivity and signal persistency allow to estimate the atmospheric stability (e.g., GAGE and GREEN, 1979; GREEN and GAGE, 1980; ROTTGER, 1980a) and to retrieve the temperature profile (e.g. GAGE and GREEN, 1982; WESTWATER et al., 1983).

An impression on the signal variability at short time and height scales can be obtained from the modified height-time intensity plot (Figure 18). This plot is obtained by suppressing the mean profile (average over many hours) and displaying only the residual power data. The thin structures, observed with vertical beam, are due to reflection from almost horizontal stratifications of refractive index changes. These are also called sheets or laminae because they are often even thinner than the normally applied range resolution of 150 m (ROTTGER and SCHMIDT, 1979). It is often found that they occur in multiples or patches which may be because they are signifying the outer boundaries of turbulence layers. They even indicate periodicities which can point out that they are connected with atmospheric wave structures. These sheets can gain a substantial increase in reflectivity by an order of magnitude or more. The fine-scale time variation of the sheets is characterized by coherence times of some seconds to some ten seconds.

The long-term variation of echoes from the mesosphere can be seen in the
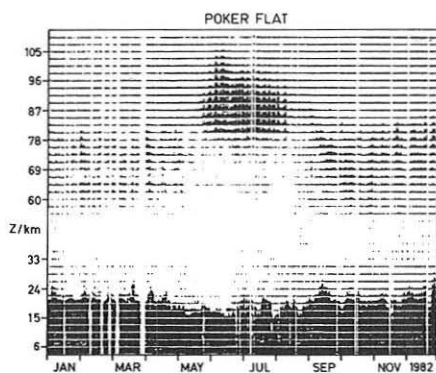

Figure 16. Occurrence of MST radar echoes (P·A $\sim 5 \cdot 10^7$ Wm^2) as a function of season (after BALSLEY et al., 1983).
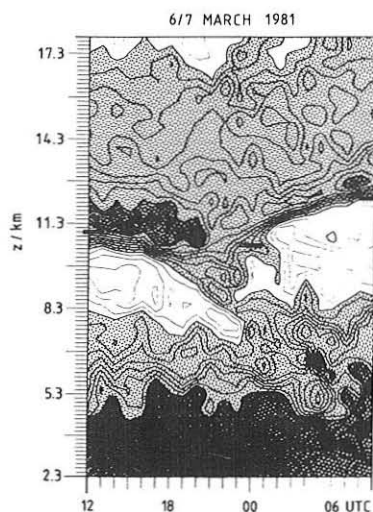
Figure 17. Reflectivity contour plot (contour line difference 2 dB, the intensity of shading corresponds to the intensity of reflectivity). The bars indicate the tropopause height measured with radiosonde (from ROTTGER and SCHMIDT, 1981).
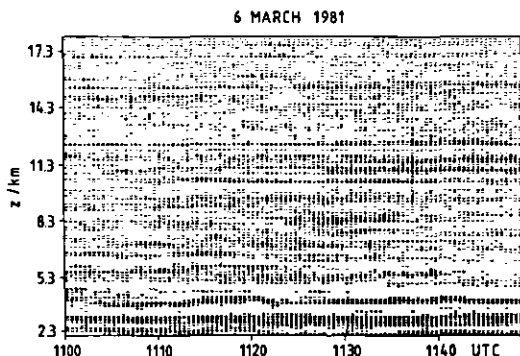
6 MARCH 1981

Figure 18. Modified height-time-intensity
plot showing the fine structure of
VHF radar returns (from HOCKING and
ROTTGER. 1983).

upper part of Figure 16. These continuous observations are from Poker Flat,
Alaska at high latitudes (BALSLEY et al., 1983), and can be used to estimate
the detectability of mesospheric echoes as compared to tropospheric/strato-
spheric echoes. It was found here that there is a very marked change in the
mesospheric echo height with season. In summer the echoes occur in the height
range 80-100 km, whereas in other seasons the echoes range from about 55 to 80
km. The transition takes place abruptly in May and September. A similar
seasonal variation of echo occurrence was also observed with the SOUSY-VHF-Radar
(CZECHOWSKY et al. 1979). ECKLUND and BALSLEY (1981) reported that the peak
signal-to-noise ratio is 1-2 orders of magnitude larger in summer than in
autumn, and it even may occur that echoes are almost absent from January to
March. It is assumed that the seasonal variation of the mesospheric echoes is
possibly connected to the variable strength of turbulence. On the other hand,
vertical gradients of electron density must be present. This connects the
echo appearance to times when either high energetic particle precipitation en-
hances the electron density in the auroral mesosphere, or normally to the day-
light hours. This diurnal variation is clearly depicted in Figure 19, which
shows the mesospheric echo appearance for low latitudes (Arecibo, Puerto Rico).
The mesospheric echo structures are also characterized by layers and sheets
with different thicknesses and lifetimes (e.g., CZECHOWSKY et al., 1979;
ROTTGER, 1980a). This makes it even more difficult to obtain a continuous alti-
tude profile of echo power and wind velocity in the mesosphere. Additionally,
meteor echoes can be observed between 90 and 110 km. These peak in the morning
hours and also were used to measure wind velocitites (e.g. AVERY et al., 1983).
The time and height variability of mesospheric echoes makes it almost impos-
sible at this time to predict the sensitivity of VHF radars for operational
studies of the mesosphere. A crude estimate can only be given (e.g. Figure 2),
and we assume a power aperture product of $10^7 - 10^8$ Wm^2 to be necessary
for suitable mesospheric investigations.

10. CONCLUSION

During recent years the number of MST radars has considerably increased,
which proves the confidence in the application of this technique for remote
sensing of the atmosphere. In Table 1 a summary of those radars is given which
are operational, under construction or in a definite planning state. Also the
incoherent scatter radars (IS) are included since all of them had been or will
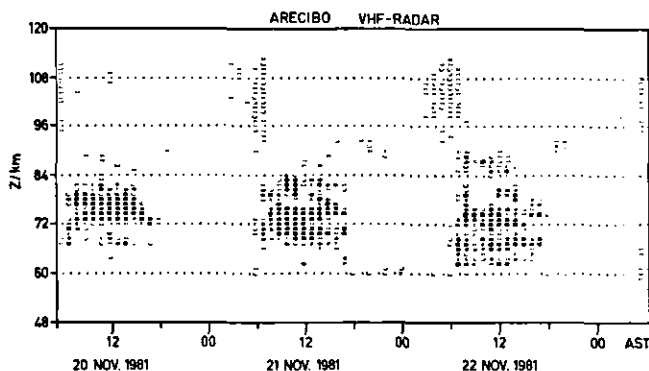
Figure 19. Occurrence of mesosphere echoes measured
with the SOUSY VHF Radar at the Arecibo Observatory
$(P \cdot A = 2.5 \cdot 10^8 Wm^2)$.

be used for ST work. Except of the Jicamarca radar, all IS radars operate at
high VHF or UHF, and are not regular MST radars in terms of the definition of
mesosphere-stratosphere-troposphere turbulence scatter. Some numbers of the
technical parameters in this list (e.g., average power, min. pulse width, duty
cycle, aperture and beam width) are uncertain to some extent, because either
varying operation modes are used or these terms are defined in a slightly dif-
ferent manner according to the different experimenters' requirements. However,
they give a gross overview and allow the determination of the minimum, average
and maximum technical standard of the transmitter-receiver-antenna system. All
these radars apply digital control, signal acquisition and processing. The
radar controller is mostly a programmable hardware unit, whereas preprocessor
(integrator, decoder, correlator, array processor) are either separate hardware
units or implemented in the supervising host computer. The application of
coding is not yet general standard but its implementation is often planned.
The smallest radars, particularly those of the Platteville-type, operate only
at a power-aperture product of $10^6 Wm^2$ and detect echoes up to the lower
stratosphere. The Jicamarca radar has the highest power-aperture product of
$10^{10} Wm^2$, but its outstanding sensitivity is still barely sufficient to de-
tect evaluable echoes from the gap region between 45 and 55 km altitude. A
large international MST system is planned for operation close to the equator
especially to allow the measurements of (vertical) velocities through almost
the entire middle atmosphere (BALSLEY, private communication). This venture is
of eminent interest for studying the dynamics of the middle atmosphere, partic-
ularly the global circulation pattern (e.g. GELLER, 1979; GAGE and VANZANDT,
1981). One of the essential contributions of MST radars to dynamical meteorol-
ogy of the middle and lower atmosphere is their capability to measure the ver-
tical velocities (e.g., GAGE 1983). The determination of instrumental limita-
tions and the development of optimum and error-minimized methods to measure the
vertical velocity is still one of the outstanding tasks for the MST radar ex-
perimenters.

    The evident capabilities of MST radars to investigate and monitor waves
and turbulence and their mutual coupling with the mean flow of the general cir-
culation pattern as well as the impact on vertical transport is recognized and
accepted (e.g., GAGE and CLARK, 1978; GAGE, 1979; WOODMAN, 1981; KLOSTERMEYER,
1981; LINDZEN, 1981; WOODMAN et al., 1981; LARSEN et al., 1982; FRITTS et al.,
1984). It is good exercise for the experimenters to fully develop existing as
well as introduce new methods to measure with optimum accuracy as many

Table I   OPERATIONAL AND PLANNED MST/ST/IS RADARS (NOVEMBER 1983)

facility	mode	lat./long. deg.	freq. MHz	ave.power kW	min.pulse width μs	duty cycle (max.)%	aperture (eff.)m²	beam width	configuration	mode	steerability	status	ref.
Arecibo/Puerto Rico	IS,ST	19N,67W	430	120	1	6	50000	0.17°	circ.dish	DB	20°(m)	op	(1)
" / "	MST		46.8	1	1	2	50000	1.7°	"	DB	"	cs	
Buckland Park/Australia	(M)ST	35S,138E	54.1	0.4	7	0.7	7500	3.2°	PAC,PAY	DB,SA	15°(1)	op	(2)
Chung-Li/Taiwan	(M)ST	25N,121E	52	4	1	2	2500	5°	PAY	DB,SA	15°(2)	cs	(3)
EISCAT/North Scandinavia	IS,ST	70N (67N)	933.5	250	2	12.5	520	0.6°	3 circ.dishes	DB	80°(m),tristat.	op	(4a)
	IS,ST	19E (27E)	224	600	2	12.5	3300	0.6°,1.7°	cyl.dish	DB	30°,60°,20°(m)	cs	(4b)
Equatorial Pacific	ST (3x)	(0,150E)	49.8	0.2	5	(2)	(5000)	(5°)	PAC	DB	15°(2)	pl/cs	(5)
India	MST	—	45-55	60	1	2.5	20000	3°	PAC	DB	20°(2)	pl	(6)
Jicamarca/Peru	MST,IS	12S,72W	49.9	200	(1)	5	80000	1.06	PAC	DB,SA	3°(m)	op	(7)
LSEET/France	ST	43N,5E	47.8	1	2	1.7	3×2000	5°	PAC	DB	15°(2)	cs	(8)
Millstone Hill/USA	IS,ST	43N,72W	440	30	2	1.6	1640	1°	circ.dish	DB	<80°(m)	op	(9)
MU/Japan	MST,(IS)	35N,136E	46.5	50	1	5	8330	3.6°	PAY	DB,SA	30°(m)	(op)	(10)
Penn.State/USA	ST (3x)	41N,78W	48-50	1	4	2	2500	5°	PAC	DB	15°(2)	pl	(11)
Platteville/USA	ST (4x)	40N,105W	49.8	1	4	1.7	2000	5°	PAC	DB	15°(2)	op/rt	(12)
Poker Flat/USA	MST	65N,147W	49.9	128	2	2	40000	1.4°	PAC	DB	15°(2)	op/rt	(13)
PROUST/France	ST	45N,2E	935	(~10)	4	(0.2)	2000 (95)		2 dishes	DB	bistatic	(pl)	(14)
Sondre Stromfjord/Greenland	IS,ST	67N,51W	1290	100	(1)	3	420	0.5°	circ.dish	DB	90°(m)	op	(15)
SOUSY/Germany	MST	52N,10E	53.5	24	1	4	3200	5°	PAY	DB,SA	12.5°(m)	op	(16a)
" /Norway	MST	69N,16E	53.5	8	1	4	8800	3°	PAY	DB	4.0°(2),5.6°	op	(16b)
Sunset/USA	ST	40N,106W	40.5	16	1	2.5 (16)	2200	4.4°,4.8°	PAC	DB	60°(2,m)	op	(17)
United Kingdom	MST	—	~50	12	1	6	5200	3.6°	PAY	DB	5°,10°(2)	pl	(18)
Urbana/USA	MST	40N,88W	40.9	6	10	1	2000	(<3°)	PAO	DB	1.5°(1),2.5°(2)	op	(19)

mode: IS = incoherent scatter (thermosphere, may include mesosphere)
         MST = mesosphere,stratosphere,troposphere
         ST = stratosphere,troposphere

antenna:
configuration: PA = phased array
               PAC = phased array, coaxial-collinear dipoles
               PAO = phased array, dipole
               PAY = phased array, Yagi
mode: DB = Doppler beam swinging
               SA = spaced antenna (interferometer)
steerability: 15°(2) = zenith angle 15° in 2 orthogonal planes (and zenith)
              15°(m) = multiple position out to 15° zenith angle (and zenith)

status: op = operational
         op/rt = routine operation (continuous)
         cs = under construction
         pl = planned

ref.: (1) Woodman+; (2) Vincent et al., 1982; (3) Brosnahan et al.+;
       (4a) Röttger et al.+; (4b) Hagfors et al., 1982; (5) Balsley+;
       (6) Koshy+; (7) Woodman and Guillén, 1974, Woodman and Farley+;
       (8) Crochet+; (9) Rastogi+; (10) Kato+; (11) Peters+;
       (12) Strauch et al., 1983, Strauch+; (13) Balsley et al., 1980,
       Balsley et al.+; (14) Glass+; (15) Watkins+; (16a) Röttger et al.,
       1978; (16b) Czechowsky et al.+; (17) Green et al., 1975, Green+;
       (18) Hall+; (19) Råyrvik and Goss+

       + = in: Bowhill (1984), Handbook for MAP 9.

parameters of waves and turbulence as possible.

Finally, it shall be stressed here that the MST radar technique recently has developed into a state which will allow its routine application for operational meteorology (e.g., ROTTGER, 1981a; BALSLEY and GAGE, 1982; HOGG et al., 1983; STRAUCH et al., 1983 and WESTWATER et al., 1983). The capabilities of the MST radars to observe continuously the 3-dimensional wind field, frontal structures, the tropopause and static stability is more and more recognized (e.g., LHERMITTE, 1979; LARSEN and ROTTGER, 1982; LARSEN, 1983b). The conclusive step-forward for the experimenter and system designer is to refine further research work in order to prepare MST radar equipment which operates continuously and unattendedly as well as to allow real-time analysis, access and transfer of relevant meteorological data.

Besides continuing to be used as very suitable research tools it now can be foreseen that within the next decade MST and ST radars may be linked in routine networks for real-time meteorological operations.

REFERENCES

Avery, S. K., A. C. Riddle and B. B. Balsley (1983), The Poker Flat radar as a meteor radar. Radio Sci., 18, 1021-1027.
Balsley, B. B. (1978a), The use of sensitive coherent radars to examine atmospheric parameters in the height range 1-100 km. Preprint Vol. 18th Conf. on Radar Meteorol., 190-193 (publ. by Am. Meteorol. Soc., Boston, MA).
Balsley, B. B. (1978b), Design considerations for coherent radar systems for probing the troposphere, stratosphere, and mesosphere. Preprint Vol. 18th Conf. on Radar Meteorol., 387-390, (publ. by Am. Meteorol. Soc., Boston, MA).
Balsley, B. B. and W. L. Ecklund (1972), A portable coaxial collinear antenna, IEEE Trans. Antennas Propagat. AP-20, 513-516.
Balsley, B. B., W. L. Ecklund, D. A. Carter and P. E. Johnston (1980), The MST radar at Poker Flat, Alaska. Radio Sci., 15, 213-223.
Balsley, B. B. and K. S. Gage (1980), The MST radar technique: Potential for middle atmospheric studies, J. Pure Appl. Geophys., 118, 452-493.
Balsley, B. B. and K. S. Gage (1981), On the vertical-incidence VHF backscattered power profile from the stratosphere, Geophys. Res. Lett., 8, 1173-1175.
Balsley, B. B. and K. S. Gage (1982), On the use of radars for operational wind profiling, Bull. Am. Meteorol. Soc., 63, 1009-1018.
Balsley, B. B., W. L. Ecklund, D. A. Carter (1983), Capabilities and limitations of the Poker Flat MST radar, HANDBOOK for MAP Vol. 9, 330-337, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana.
Battan, L. J. (1973), Radar Observation of the Atmosphere, The Univ. of Chicago Press.
Bolgiano, R., Jr. (1968), The general theory of turbulence - Turbulence in the atmosphere, in: Winds and Turbulence in Stratosphere, Mesosphere and Ionosphere, edited by K. Rawer, 371-400, North-Holland Publ. Co., Amsterdam.
Bowhill, S. A. and B. Edwards (Editors) (1983), Handbook for MAP, Vol. 9, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana.
Bowhill, S. A. and B. Edwards (Editors) (1984), Handbook for MAP, Vol. 14, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana, in press.
Bowles, K. L. (1958), Observation of vertical-incidence scatter from the ionosphere at 41 Mc/sec, Phys. Rev. Lett., 1, 454-455.
Briggs, B. H. (1977), The analysis of moving patterns by correlation methods, Rep. ADP 148, Dep. Phys., Univ. of Adelaide, Australia.
Briggs, B. H. (1980), Radar observations of atmospheric winds and turbulence: a comparison of techniques, J. Atmos. Terr. Phys., 42, 823-833.

228

Carter, D. A., B. B. Balsley and W. L. Ecklund (1980), The Poker Flat MST ra-
    dar: signal analysis and data processing techniques with examples, Preprint
    Vol. 19th Conf. on Radar Meteorol., 563-567 (publ. by Am. Meteorol. Soc.,
    Boston, MA).
Clark, W. L. and D. A. Carter (1980), Real-time scaling of atmospheric para-
    meters from radars using the MST technique, Preprint Vol. 19th Conf. on
    Radar Meteorol., 599-604, (publ. by Am. Meteor. Soc., Boston, MA).
Crane, R. K. (1980), A review of radar observations of turbulence in the lower
    stratosphere, Radio Sci., 15, 177-193.
Czechowsky, P. and K. Meyer (1980), Das Antennensystem des SOUSY-VHF-Radars,
    Rep. MPAE-T-00-80-19, Max-Planck-Institut fur Aeronomie, Katlenburg-
    Lindau, Germany.
Czechowsky, P., J. Klostermeyer, J. Rottger, R. Ruster, G. Schmidt and R. F.
    Woodman (1976), The SOUSY-VHF-Radar for tropospheric, stratospheric and
    mesospheric sounding, Preprint Vol. 17th Conf. on Radar Meteorol., 349-353,
    (publ. by Am. Meteorol. Soc., Boston, MA).
Czechowsky, P., R. Ruster and G. Schmidt (1979), Variations of mesospheric
    structures in different seasons, Geophys. Res. Lett., 6, 459-462.
Czechowsky, P., G. Schmidt and R. Ruster (1984), The mobile SOUSY Doppler ra-
    dar - Technical design and first results, Radio Sci., 19, 441-450.
Doviak, R. J. and D. S. Zrnic' (1984), Doppler Radar and Weather Observations,
    Academic Press, Inc., Orlando.
Ecklund, W. L. and B. B. Balsley (1981), Long-term observations of the arctic
    mesosphere with the MST radar at Poker Flat, Alaska, J. Geophys. Res.,
    86, 7775-7780.
Farley, D. T., H. M. Ierkic and B. G. Fejer (1981), Radar interferometry: A
    new technique for studying plasma turbulence in the ionosphere, J. Geophys.
    Res., 86, 1467-1472.
Flock, W. L. and B. B. Balsley (1967), VHF radar returns from the D region of
    the equatorial ionosphere, J. Geophys. Res., 72, 5537-5541.
Fritts, D. C., M. A. Geller, B. B. Balsley, M. L. Chanin, I. Hirota, J. R.
    Holton, S. Kato, R. S. Lindzen, M. R. Schoeberl, R. A. Vincent and R. F.
    Woodman (1984), Research status and recommendations from the Alaska Workshop
    on Gravity Waves and Turbulence in the Middle Atmosphere, Bull. Am.
    Meteorol. Soc., 65, 149-159.
Fukao, S., T. Sato, S. Kato, R. M. Harper, R. F. Woodman and W. E. Gordon
    (1979), Mesospheric winds and waves over Jicamarca on May 23-24, 1974,
    J. Geophys. Res., 84, 4379-4386.
Fukao, S., S. Kato, T. Aso, M. Sasada and T. Makihira (1980), Middle and upper
    atmosphere radar (MUR) under design in Japan, Radio Sci., 15, 225-231.
Fukao, S., T. Sato, N. Yamasaki, R. M. Harper and S. Kato (1982), Winds measured
    by a UHF Doppler radar and rawinsondes: Comparisons made on twenty-six days
    (August-September 1977) at Arecibo, Puerto Rico, J. Appl. Meteor., 21,
    1357-1363.
Gage, K. S. (1979), Evidence for a $k^{-5/3}$ law inertial range in mesoscale two-
    dimensional turbulence, J. Atmos. Sci., 36, 1950-1954.
Gage, K. S. (1983), On the measurement of vertical velocity by MST radar, Hand-
    book for MAP, Vol. 9, 215-226, SCOSTEP Secretariat, Dep. Elec. Eng., Univ.
    IL, Urbana.
Gage, K. S. and B. B. Balsley (1978), Doppler radar probing of the clear atmo-
    sphere, Bull. Am. Meteorol. Soc., 59, 1074-1093.
Gage, K. S. and B. B. Balsley (1980), On the scattering and reflection mecha-
    nisms contributing to clear air radar echoes from the troposphere, strato-
    sphere and mesosphere, Radio Sci., 15, 243-257.
Gage, K. S. and W. L. Clark (1978), Mesoscale variability of jet stream winds
    observed by the Sunset VHF Doppler radar, J. Appl. Meteorol., 17, 1412-
    1416.
Gage, K. S. and J. L. Green (1978), Evidence for specular reflection from
    monostatic VHF radar observations of the stratosphere, Radio Sci., 13,
    991-1001.

Gage, K. S. and J. L. Green (1979), Tropopause detection by partial specular reflection with very-high-frequency radar, Science, 203, 1238-1240.

Gage, K. S. and J. L. Green (1982), A technique for determining the temperature profile from VHF radar observations, J. Appl. Meteor., 21, 1146-1149.

Gage, K. S. and T. E. VanZandt (1981), Wind measurement techniques available for the middle atmosphere program, J. Geophys. Res., 86, 9591-9598.

Gage, K. S., D. A. Carter and W. L. Ecklund (1981), The effect of gravity waves on specular echoes observed by the Poker Flat MST radar, Geophys. Res. Lett., 8, 599-602.

Geller, M. A. (1979), Dynamics of the middle atmosphere, J. Atmos. Terr. Phys., 41, 683-705

Gossard, E. E. and R. G. Strauch (1983), Radar Observation of Clear Air and Clouds, Elsevier, Amsterdam.

Green, J. L. and K. S. Gage (1980), Observations of stable layers in the troposphere and stratosphere using VHF radar, Radio Sci., 15, 395-405.

Green, J. L., J. M. Warnock, R. H. Winkler and T. E. VanZandt (1975), A sensitive VHF radar for the study of winds, waves and turbulence in the troposphere, stratosphere and mesosphere, Preprint Vol. 16th Conf. on Radar Meteorol., 313-315 (publ. by Am. Meteorol. Soc., Boston, MA).

Hagfors, T., P. S. Kildal, H. J. Karcher, B. Liesenkotter and G. Schroer (1982), VHF parabolic cylinder antenna for incoherent scatter radar research, Radio Sci., 17, 1607-1621.

Hardy, K. R. (1972), Studies of the clear atmosphere using high power radar in: Remote Sensing of the Troposphere (edit. V. E. Derr), chapter 14, Rep. of Wave Propagat. Lab., ERL, Boulder, CO, issued August 15, 1972.

Harper, R. M. and W. E. Gordon (1980), A review of radar studies of the middle atmosphere, Radio Sci., 15, 195-211.

Hocking, W. K. (1983a), On the extraction of atmospheric turbulence parameters from radar backscatter Doppler spectra – I, Theory, J. Atmos. Terr. Phys., 45, 89-102.

Hocking, W. K. (1983b), Mesospheric turbulence intensitites measured with a HF radar at 35°S – II, J. Atmos. Terr. Phys., 45, 103-114.

Hocking, W. K. (1983c), The spaced antenna drift method, HANDBOOK for MAP, Vol. 9, 171-186, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana.

Hocking, W. K. and J. Rottger (1983), Pulse length dependence of radar signal strengths for Fresnel backscatter, Radio Sci., 18, 1312-1324.

Hogg, D. C., M. T. Decker, F. D. Guiraud, K. B. Earnshaw, D. A. Merritt, K. P. Moran, W. B. Sweezy, R. G. Strauch, E. R. Westwater and C. G. Little (1983), An automatic profiler of the temperature, wind and humidity in the troposphere, J. Climate Appl. Meteorol., 22, 807-831.

Ierkic, H. M. and J. Rottger (1984), Mesospheric measurements of irregularity patches using a 3 antenna interferometer, Handbook for MAP, Vol 14, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana (in press).

James, P. K. (1980), A review of radar observations of the troposphere in clear air conditions, Radio Sci., 15, 151-175.

Klostermeyer, J. (1981), MST radars: Advanced tools for gravity wave studies, Nature, 292, 107-108.

Koscielny, A. J. and R. J. Doviak (1983), An evaluation of the accuracy of some radar wind profiling techniques, HANDBOOK for MAP, Vol. 9, 192-209, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana.

Larsen, M. F. (1983a), The MST radar technique: A tool for investigations of turbulence spectra, HANDBOOK for MAP, Vol. 9, 250-255, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana.

Larsen, M. F. (1983b), The MST radar technique: Requirements for operational weather forecasting, HANDBOOK for MAP, Vol. 9, 3-11, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana.

Larsen, M. F. (1983c), Can a VHF Doppler radar provide synoptic wind data? Preprint Vol. 5th Symp. on Meteorol. Observations and Instrumentation, 183-190, (publ. by Am. Meteorol. Soc., Boston, MA).

Larsen, M. F. and J. Rottger (1982), VHF and UHF Doppler radars as tools for

synoptic research, Bull. Am. Meteorol. Soc., 63, 996-1008.

Larsen, M. F., M. C. Kelley and K. S. Gage (1982), Turbulence spectra in the upper troposphere and lower stratosphere at periods between 2 hours and 40 days, J. Atmos. Sci., 39, 1035-1041.

Lhermitte, R. (1979), Advancements in remote sensing of the atmosphere, Rev. Geophys. Space Phy., 17, 1833-1840.

Lindzen, R. S. (1981), Turbulence and stress owing to gravity wave and tidal breakdown, J. Geophys. Res., 86, 9707-9714.

NAIC (1976), The Users Manual for the Arecibo Observatory. National Astronomy and Ionosphere Center. Arecibo Observatory, Arecibo, Puerto Rico, September 1976.

Nastrom, G. D., K. S. Gage and B. B. Balsley (1982), Variability of C_n^2 at Poker Flat, Alaska, from mesosphere, stratosphere, troposphere (MST) Doppler radar observations, Optical Engineering, 21, 347-351.

Ochs, G. R. (1965), The large 50 Mc/s dipole array at Jicamarca radar observatory, NBS Rep. 8772, National Bureau of Standards, U.S. Dept. of Commerce, Boulder, Colorado.

Rastogi, P. K. (1981), Radar studies of gravity waves and tides in the middle atmosphere: a review, J. Atmos. Terr. Phys., 43, 511-524.

Rastogi, P. K. (1983), Data processing techniques used with MST radars - a review, HANDBOOK for MAP, Vol. 9, 477-488, SCOSTEP Secretariat, Dep. Elec. Eng., Univ. IL, Urbana.

Rastogi, P. K. and O. Holt (1981), On detecting reflections in presence of scattering from amplitued statistics with application to D region partial reflection, Radio Sci., 16, 1431-1443.

Rastogi, P. K. and J. Rottger (1982), VHF radar observations of coherent reflections in the vicinity of the tropopause, J. Atmos. Terr. Phys., 44, 461-469.

Rastogi, P. K. and R. F. Woodman (1974), Mesospheric studies using the Jicamarca incoherent-scatter radar, J. Atmos. Terr. Phys., 36, 1217-1231.

Rottger, J. (1980a), Structure and dynamics of the stratosphere and mesosphere revealed by VHF radar investigations, J. Pure Appl. Geophys., 118, 494-527.

Rottger, J. (1980b), Reflection and scattering of VHF radar signals from atmospheric refractivity structures, Radio Sci., 15, 259-276.

Rottger, J. (1980b), Reflection and scattering of VHF radar signals from atmospheric refractivity structures, Radio Sci., 15, 259-276.

Rottger, J. (1981a), The capabilities of VHF radars for meteorological observations, ESA SP-165, 143-148 (publ. by European Space Agency, Paris).

Rottger, J. (1981b), Wind variability in the stratosphere deduced from spaced antenna VHF radar measurements, Preprint Vol. 20th Conf. on Radar Meteorol., 22-29 (publ. by Am. Meteorol. Soc., Boston MA).

Rottger, J. (1981c), Investigations of lower and middle atmosphere dynamics with spaced antenna drifts radars, J. Atmos. Terr. Phys., 43, 277-292.

Rottger, J. (1984), The MST radar technique, compared with the Thomson scatter radar technique - A comparison of coherent and incoherent scatter radars, EISCAT Techn. Rpt., EISCAT Scient. Association, Kiruna, Sweden (in preparation).

Rottger, J. and P. Czechowsky (1980), Tropospheric and stratospheric wind measurements with the spaced antenna drifts technique and the Doppler beam swinging technique using a VHF radar, Preprint Vol. 19th Conf. on Radar Meteorol., 577-584, (publ. by Am. Meteorol. Soc., Boston, MA).

Rottger, J. and C. H. Liu (1978), Partial reflection and scattering of VHF radar signals from the clear atmosphere, Geophys. Res. Lett., 5, 357-360.

Rottger, J. and G. Schmidt (1979), High-resolution VHF radar sounding of the troposphere and stratosphere, IEEE Trans. Geosci. Electr., GE-17, 182-189.

Rottger, J. and G. S. Schmidt (1981), Characteristics of frontal zones deter-

mined from spaced antenna VHF radar observations, <u>Preprint Vol. 20th Conf.</u> <u>on Radar Meteorol.</u>, 30-37, (publ. by Am. Meteorol. Soc., Boston, MA).

Rottger, J. and R. A. Vincent (1978), VHF radar studies of tropospheric velocities and irregularities using spaced antenna techniques, <u>Geophys. Res.</u> <u>Lett.</u>, <u>5</u>, 917-920.

Rottger, J., J. Klostermeyer, P. Czechowsky, R. Ruster and G. Schmidt (1978), Remote sensing of the atmosphere by VHF radar experiments, <u>Naturwissen-</u> <u>schaften</u>, <u>65</u>, 285-296.

Rottger, J., P. Czechowsky and G. Schmidt (1981), First low-power VHF radar observations of tropospheric, stratospheric and mesospheric winds and turbulence at the Arecibo Observatory, <u>J. Atmos. Terr. Phys.</u>, <u>43</u>, 789-800.

Royrvik, O. (1983a), Spaced antenna drift at Jicamarca, mesospheric measurements, <u>Radio Sci.</u>, <u>18</u>, 461-476.

Royrvik, O. (1983b), VHF radar signals scattered from the equatorial mesosphere, <u>Radio Sci.</u>, <u>18</u>, 1325-1335.

Ruster, R., P. Czechowsky and G. Schmidt (1980), VHF-radar measurements of dynamical processes in the stratosphere and mesosphere, <u>Geophys. Res. Lett.</u>, <u>7</u>, 999-1002.

Ruster, R. and R. F. Woodman (1978), Digital filtering, calibration and correlation analysis of radar-echoes from the tropo- and stratosphere, <u>Kleinheu-</u> <u>bacher Berichte</u>, <u>21</u>, 239-246, (publ. by Fernmeldetechnisches Zentralamt, Darmstadt, Germany).

Sato, T. and S. Fukao (1982), Altitude smearing due to instrumental resolution in MST radar measurements, <u>Geophys. Res. Lett.</u>, <u>9</u>, 72-75.

Sato, T. and R. F. Woodman (1980), Spectral parameter estimation of CAT radar echoes in the presence of fading clutter, Preprint Vol. 19th Conf. on Radar *Meteorol.*, 568-574 (publ. by Am. Meteorol. Soc., Boston, MA).

Schmidt, G., R. Ruster and P. Czechowsky (1979), Complementary code and digital filtering for detection of weak VHF radar signals from the mesosphere, <u>IEEE</u> <u>Trans. Geosci. Electr.</u>, <u>GE-17</u>, 154-161.

Skolnik, M. I. (1970) (Editor), <u>Radar Handbook</u>, McGraw-Hill, Inc., New York.

Strauch, R. G., D. A. Merritt, K. P. Moran, K. B. Earnshaw and D. van de Kamp (1983), Tropospheric wind profiling with Doppler radar. Preprint Vol. 21th *Conf. on Radar Meteorol.*, 118-125, (publ. by Am. Meteorol. Soc., Boston, MA).

VanZandt, T. E., K. S. Gage and J. M. Warnock (1981), An improved model for the calculation of profiles of $C_n{}^2$ and ϵ in the free atmosphere from background profiles of wind, temperature and humidity, Preprint Vol. 20th Conf. on Radar Meteorol., 129-135, (publ. by Am. Meteorol. Soc., Boston, MA).

VanZandt, T. E. and R. A. Vincent (1983), Is VHF Fresnel reflectivity due to low frequency buoyancy waves? Preprint Vol. 21st Conf. on Radar Meteorol., 126-128, (publ. by Am. Meteorol. Soc., Boston, MA).

Vincent, R. A. and I. M. Reid (1983), HF Doppler measurements of mesospheric gravity wave momentum fluxes, <u>J. Atmos. Sci.</u>, <u>40</u>, 1321-1333.

Vincent, R. A. and J. Rottger (1980), Spaced antenna VHF radar observations of tropospheric velocities and irregularities, <u>Radio Sci.</u>, <u>15</u>, 319-335.

Vincent, R. A., W. G. Elford and B. H. Briggs (1982), A VHF radar for atmospheric studies, <u>The Australian Physicist</u>, <u>19</u>, 70-73.

Walker, J. C. G. (1979), Radar measurement of the upper atmosphere, <u>Science</u>, <u>206</u>, 180-189.

Westwater, E. R., M. T. Decker, A. Zachs and K. S. Gage (1983), Ground-based remote sensing of temperature profiles by a combination of microwave radiometry and radar, <u>J. Climate Appl. Meteorol.</u>, <u>22</u>, 126-133.

Wilson, D. A. and L. J. Miller (1972), Atmospheric motion by Doppler radar in: Remote Sensing of the Troposphere (edit. V. E. Derr), chapter 13 (Rep. of Wave Propag. Lab., ERL, Boulder, CO, issued August 15, 1972).

Wilson, J., R. Carbone, H. Baynton and R. Serafin (1980), Operational application of meteorological Doppler radar, <u>Bull. Am. Meteorol. Soc.</u>, <u>61</u>, 1154-1168.

Woodman, R. F. (1981), Turbulence in the middle atmosphere: A review, <u>Handbook</u>

for MAP, Vol. 2, (edit. S. K. Avery), 293-300, SCOSTEP Secretariat, Univ. IL, Urbana.

Woodman, R. F. and A. Guillen (1974), Radar observations of winds and turbulence in the stratosphere and mesosphere, J. Atmos. Sci., 31, 493-505.

Woodman, R. F. and T. Hagfors (1969), Methods for the measurement of vertical ionospheric motions near the magnetic equator by incoherent scattering, J. Geophys. Res., 74, 1205-1212.

Woodman, R. F., R. P. Kugel and J. Rottger (1980), A coherent integrator-decoder preprocessor for the SOUSY-VHF-Radar. Radio Sci., 15, 233-242.

Woodman, R. F., P. K. Rastogi and T. Sato (1981), Evaluation of effective eddy diffusive coefficients using radar observations of turbulence in the stratosphere, Handbook for MAP, Vol. 2, (edit. S. K. Avery), 363-369, SCOSTEP Secretariat, Dep. Elec. Engin., Univ. IL, Urbana.

15. PARTIAL REFLECTION SPACED ANTENNA WIND MEASUREMENTS

G. J. Fraser

Physics Department
University of Canterbury
Christchurch, New Zealand

INTRODUCTION

The partial reflection (PR) radars discussed here use frequencies in the 2-6 MHz range and operate in a pulse mode at vertical incidence. The first observations of partial reflections from the lower ionosphere were reported by GARDNER and PAWSEY (1953). Their work led to further investigations at a number of locations into the morphology and the nature of the irregularities responsible for the reflections and to the development of techniques for measuring D-region electron densities by the so-called differential absorption method (see Manson, this volume).

The characteristics of the partial reflections (PR) as observed at a site near Christchurch, N.Z. at 44°S (GREGORY, 1956, 1961) suggested that they would be suitable for the measurements of mesospheric winds using the spaced antenna method which was originally applied to total reflections by for example, MITRA (1949). Subsequent modification of the Christchurch PR installation allowed wind measurements using this technique and the first results were reported by FRASER (1965, 1968). The method was then applied at other mid-latitude locations including Adelaide at 35°S (STUBBS, 1973) and at Saskatoon at 52°N (GREGORY and REES, 1971) and later at tropical (VINCENT and BALL, 1981) and polar latitudes (FRASER, 1983).

Although MF/HF radars are generically referred to as "partial reflection" the actual mechanisms causing the echoes range from reflections from sharply bounded irregularities in refractive index to scattering from quasi-isotropic irregularities whose scales are the order of half the radar wavelength. Echoes are obtained from minimum heights as low as 50-60 km, depending on system sensitivity, up to a height where the wave is totally reflected, in the E or F region. The variations in refractive index are predominantly in the vertical but there are also significant variations in the horizontal direction. The resultant horizontal variations modulate the incident wavefront so that the reflected or scattered wave is equivalent to an angular spectrum of plane waves. It is the interference of the angular spectrum components which form the Fresnel diffraction pattern on the ground and the motion of the reflecting regions causes motion of the diffraction pattern with twice the velocity. The factor of two arises from the diffraction geometry and is associated with the so-called "point source" effect (FELGATE, 1970; BRIGGS, 1980). As the pattern moves past a stationary antenna the induced e.m.f. varies in amplitude (radiowave fading). A comparison of the fading recorded at a minimum of three spaced antennas on the ground gives an estimate of the horizontal pattern velocity and hence the horizontal velocity of the reflecting region. More details of the data analysis techniques used are given by Briggs in an accompanying article. Here we concentrate on the experimental techniques.

THE NATURE OF PARTIAL REFLECTIONS

In order to discuss the techniques used in the PR spaced antenna method it is necessary to review briefly the nature of the partially reflecting irregularities as they determine the system parameters.

The partially reflecting regions display spatial characteristics ranging from thin stratified layers to thick turbulent-like layers. A number of measurements show that below 80 km the reflecting regions are frequently thin (<2 km thick) with relatively large horizontal scale while above 80 km the layers are thicker (GREGORY, 1961; FRASER and VINCENT, 1970). Measurements of the angular spectra show that the regions below 80 km are quasi-specular with narrow angular spectra ($\sim 2°-10°$ from the zenith) while above 80 km the echoes are returned over a wider range of angles (up to $15°-20°$). The wide range of reflecting region characteristics implies a wide range of irregularity scale sizes which can complicate the optimum choice of antenna spacing. However, the ability of the spaced antenna method to work with small angles of scatter means that it is well suited to cope with a variety of conditions (BRIGGS, 1980).

The reflecting regions generally have an irregular structure so the diffraction pattern is also irregular. The diffraction pattern scale is usually defined by the separation between two antennas when the correlation coefficient of their fading records is 0.5. This scale is typically 1-2 radio wavelengths ($\sim 100-200$ m) for D-region reflections. The antenna spacing used is a compromise between the ambiguities which arise with increasing spacing and the errors which may arise at close spacing. GOLLEY and ROSSITER (1970, 1971) concluded that the optimum spacing is one for which the zero-lag cross-correlation is about 0.5.

While the D-region echoes usually show a random structure, fringes which are produced by interference between two or more discrete rays are occasionally observed (FELGATE and GOLLEY, 1971). If the interfering rays come from an isoionic contour perturbed by a gravity wave, it is possible that the motion measured on the ground may be related to the phase velocity of the wave rather than the motion of the neutral atmosphere. Another factor which may need to be taken into account is the effect of the geomagnetic field which will have an influence on the motion of electrons at heights where the electron-neutral collision frequency is low. This effect does not appear to be important at heights below 95-100 km (FRASER, 1968). Overall, experimental intercomparisons with a variety of techniques show that the spaced antenna method using partial reflections from the D-region is a reliable and inexpensive method for measuring winds in the mesosphere and lower thermosphere (STUBBS, 1973; VINCENT et al., 1977).

The most appropriate frequency to use depends on a number of factors including the strength of the reflections, radiowave absorption and external interference. It can be shown that the reflection coefficients which in turn determine the amplitudes of the received signals depend on electron density and on the factor $\{f(f \pm f_L)\}^{-1}$ (BUDDEN, 1961) where f is the wave frequency and f_L is the longitudinal component of the electron gyrofrequency. The $-$ sign is associated with extraordinary (X) component and the $+$ sign with the ordinary (O) component. In either case the maximum sensitivity is obtained if $f(f \pm f_L)$ is small. Typically f_L is near 1.5 MHz at mid-latitudes and most PR radars operate in the 2-3 MHz range; in particular the X-component is strongly reflected if f is close to f_L.

A disadvantage of using too low a frequency is the lowering, for a given electron density, of the height at which total reflection occurs. At frequencies of 2-3 MHz the total reflection usually comes from the normal E-region at heights near 100 km at midday, but when some types of sporadic-E ionization are present, totally reflected echoes can occur at heights as low as 90-95 km, thus obscuring the partially reflecting irregularities at those heights.

Although the X-mode reflection coefficients are stronger than those for the O-mode the X-mode signals are more heavily absorbed during the daytime than the

O-mode whose amplitude increases in an approximately exponential manner with height (BELROSE, 1970) so that there continues to be adequate signal strength for spaced antenna wind measurements. At night both components, but especially the X-mode, give significant partial reflections from heights above about 80 km; at lower heights there are usually insufficient electrons to produce usable reflections.

When linearly polarized arrays are used for transmission and reception there is the possibility of interference between the two magnetoionic components. Any difficulties due to this effect may be removed by transmitting and/or receiving one or other component. As both components are elliptically polarized this implies some complexity in the antenna arrays and instrumenta-tion. However, at mid- to high-latitudes the polarization ellipses are almost circular so circularly polarized transmitting arrays have proved succesful - for example at Adelaide and Saskatoon. For reasons discussed above the trans-missions are usually O-mode during the day and X-mode at night. Near the geomagntic equator the characteristic polarizations are of course linear. If linear arrays are used for transmission and reception (e.g., Christchurch and Scott Base) the importance of magneto-ionic interference depends on the relative amplitudes and phase differences between the components. At low heights each refractive index differs so little from unity that there is negligible difference between the component diffraction patterns. The correlation coefficient between the X and O-component reflections is usually very high (Von Biel, private communication) so that data loss due to this cause is small.

DIURNAL AND SEASONAL VARIABILITY OF MEASUREMENTS

Analyses for tidal components or the diurnal variation of gravity wave activity or diffraction pattern parameters require wind measurements spread uniformly throughout 24 hours. The major obstacle to such continuous measure-ment is the low electron density below ~80 km at night.

Figure 1 shows the day/night contrast in the number of wind observations for 6-hour periods centered on midnight and mid-day at Christchurch (44S) in summer (December 1981) and winter (June 1981). The almost complete disappearance of nighttime reflections below 82-83 km is apparent. The observation frequency for the 6-hour "twilight" intervals was very similar to the nighttime distribution. The figure shows the actual fraction of data collected - under all conditions, including lack of ionization, radio frequency interference, adjustment of the operating schedule to exclude heights below 77.5 km at night, and some overlap in the height scanning sequency at 77.5 and 90 km.

Figure 2 shows the "night" and "day" distributions at Scott Base (78S) in summer (December 1982). At Scott Base observations are mainly lost through lack of ionization as radio frequency interference and operating schedule changes are negligible.

These figures illustrate the dependence of usable data on height. The strength of the returned echoes is one factor controlling whether data is usable. At Christchurch and Scott Base no quantitative measure of echo power is recorded. However HOCKING (1981) has measured relative echo power as a function of height at Adelaide and Townsville (Figures 3 and 4). The decrease in X component with height during the day is very apparent, in contrast with the large X-component returned from the diminished electron density at night. The O-component echo power increases with height up to about 90 km in both day and night. During the day the O-component echo power increases above 90 km due to the formation of the E layer by solar radiation.
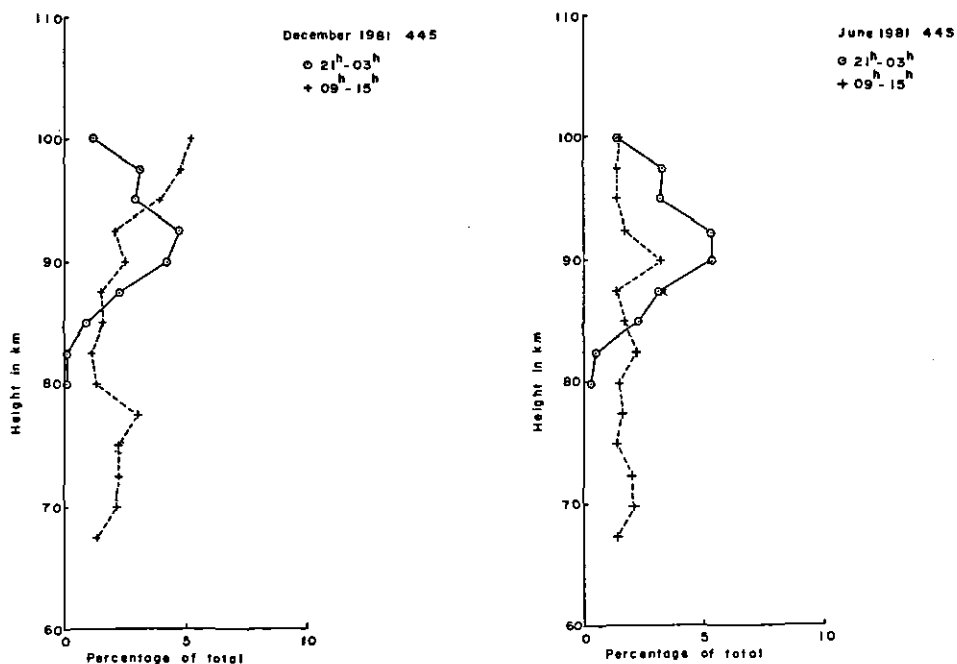
236

Figure 1. Height distribution of usable wind observations
at Christchurch (44S), given as a percentage of the
total number of usable observations. The night-time data
collection rates are similar in winter (June) and summer
(December). The day-time rates are also similar, but in
summer there is a marked increase in rate above 93 km.
The distributions for 15h – 21h and 03^h – 06^h closely
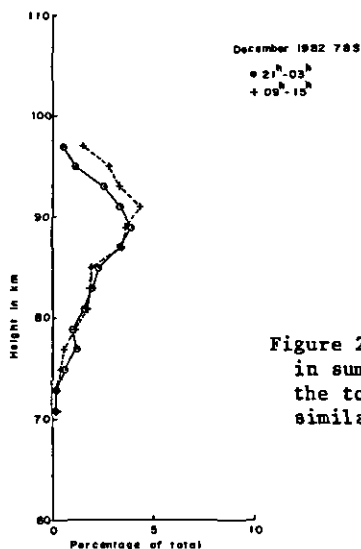resemble those for 21^h – 03^h.

Figure 2. Height distribution of usable wind observations
in summer at Scott Base (78S), given as a percentage of
the total number of usable observations. The expected
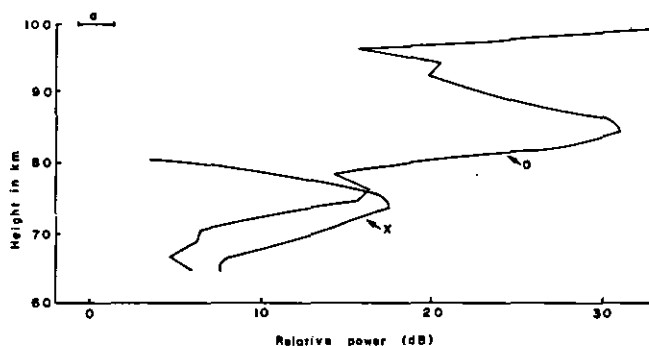similarity between "day" and "night" is evident.

Figure 3. Relative echo power as a function of height at
Adelaide (35S) (HOCKING, 1981) in day-time. The
horizontal line at 100 km indicates the typical day-
time noise level. The data are averages over 1 - 3
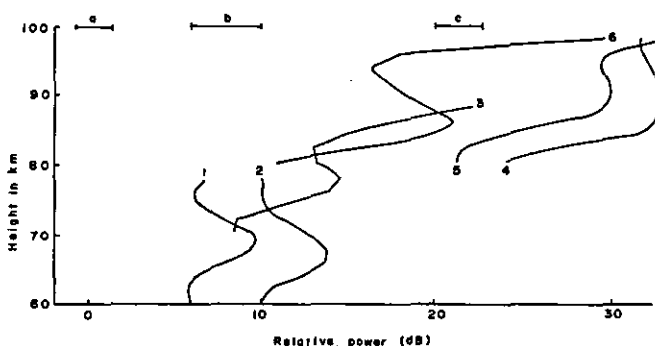minutes taken at: 1425 on 31 May 1977 (X-mode);
1224 on 20 July 1977 (O-mode).

Figure 4. Relative power of O-mode and X-mode echoes
as a function of height at Townsville (19S) (HOCKING,
1981) in early summer 1977. The horizontal lines at
the top are estimates of the radio noise: (a) during
the day after integrating 8 pulses from a coherent
detector; (b) during the day with no integration;
(c) at night with no integration. The data are 1-3
minute averages taken at: 1415 on 2 November (X-mode);
1315 on 2 November (X-mode); 2200 on 2 November (O-
mode); 2200 on 31 October (X-mode); 2340 on 31 October
(X-mode); 1400 on 1 November (O-mode). The dominance
of the X-mode at low altitudes and at night is
apparent, as are the differences between observations
only one or two hours apart.

GREGORY et al. (1982) have published a comprehensive set of relative echo power contours for Saskatoon which show in more detail the variations with local time in May and December. They also give the number of wind observations per hour as contours in height and time for winter and summer.

Echoes from altitudes as low as 40-50 km have been observed at Scott Base (78S) (GREGORY, 1983; von Biel, private communication) but they appear to be of too short a duration to permit a reliable estimate of mean or tidal winds.

RADAR SYSTEM PARAMETERS

The radar system gain is important for two reasons.

(1) The echo strength diminishes as the altitude decreases because of the decreasing electron density. The higher the system gain the lower is the minimum altitude at which echoes can be detected.

(2) Adequate resolution of tidal components requires observations over as much of the 24-hour interval as possible. At middle and low latitudes solar control of the ionization results in very weak echoes below 80 km at night (when the radio-frequency interference tends to be high) so high gain and adequate inter-ference-rejecting techniques are helpful in increasing the fraction of the day over which tidal components can be assessed.

For maximum sensitivity high transmitter pulse power (50-100 kW) and large transmitting arrays (10-15 dB gain) are desirable although good results have been obtained with small arrays or single antennas at Mawson (68S), Scott Base (78S) and Townsville (19S).

A minimum of 3 spaced receiving antennas is required for spatial sampling of the diffraction pattern. To minimise timing errors, these antennas should be arranged in an equilateral triangle (BARBER, 1957). The optimum antenna spacing for sampling the diffraction pattern is $\sim 1\lambda$ (GOLLEY and ROSSITER, 1971) so large arrays cannot be used as the individual antennas. Single dipoles are common although a pair of dipoles spaced $\lambda/2$ and fed in phase has some advantages in reducing low-angle interference (FRASER, 1968) and comprise an array which is still electrically small.

(a) Receiver Dynamic Range

Receivers for use in partial reflection wind measurements must have high sensitivity to observe the fading minima of the very weak partial reflections below 70 km. They must also amplify without saturation the much stronger partial reflections from around 90 km. An even greater dynamic range is needed if totally-reflected echoes from E or E_s layers are also made use of. At Christchurch and Saskatoon those totally-reflected signals are frequently used, but seldom at Adelaide; they do not fall within the scope of the present review, and will not be discussed further. The mean echo powers of the partial reflections (HOCKING, 1981; GREGORY et al., 1982) span a range of 40-60 dB. If we assume that the received signal has a Rayleigh distribution, the amplitude range for 99% of the time is 54 dB. Consequently a receiver dynamic range of 100-110 dB is indicated.

A maximum output of ~ 10 V is typical of solid state electronics. The lower limit of detection is about 3 mV with an active rectifier comprising an operational amplifier with diode feedback. This gives a detector with a dynamic range of 70 dB, so some external attenuation may be required.

(b) Coherent and Incoherent Detection

Incoherent detectors are non-linear devices which rectify the radio-frequency signal. At low signal-to-noise ratios the non-linearity causes cross-modulation of signal and noise components. The output signal-to-noise power ratio is low input signal-to-noise ratios (PANTER, 1965)

$$(S/N)_{out} \simeq (S/N)_{in}^{2} .$$

As the minimum detectable signal-to-noise ratio is less than 0 dB (LAWSON and UHLENBECK, 1950) the use of incoherent detection results in an output signal-to-noise ratio which is less than that at the input.

With coherent ("phase-sensitive", "synchronous", "product", etc.) detectors the input and output signal-to-noise ratios are approximately the same. For this reason the increased received complexity is justified by the increased sensitivity.

The phase ambiguity in coherent detectors is usually resolved by using two detectors whose reference inputs differ in phase by 90°, giving detector outputs which are the in-phase and quadrature components of the signal at the detector inputs. These two components may be considered as the real and imaginary components of the complex signal amplitude. The complex cross-correlation coefficient of two complex signals $S_1(t)$ and $S_2(t)$ is:

$$\rho(\tau) = \frac{<S_1(t) \; S_2^{*} \; (t + \tau)>}{[<S_1^{2}(t)> \; <S_2^{2}(t)>]^{1/2}}$$

The full correlation analysis for wind determination uses the magnitude of the complex correlation coefficient (VINCENT and BALL, 1981).

(c) Post-Detection Filtering

The maximum signal-to-noise ratio is obtained when the pre-detector passband is the Fourier transform of the pulse shape, and any kind of post-detector filter will decrease the signal-to-noise ratio (VAN VLECK and MIDDLETON, 1946). However, if the echo pulse has a constant amplitude, an improvement in signal-to-noise ratio is obtained by adding (integrating) N successive pulses. The improvement in the amplitude signal-to-noise ratio is the usual $N^{1/2}$ when adding the output pulses from a coherent detector, but only $N^{1/4}$ when incoherent detection is used (LAWSON and UHLENBECK, 1950).

If the amplitude of the echo pulse is not constant, the integration time must be restricted so that the fading of the echo pulse is not smoothed excessively. Simple addition of N pulses separated in time by N t is equivalent to a low-pass filter with a response proportional to

$$\frac{\sin (N\pi \; f\Delta t)}{(N \; f \; t)}$$

The FWHP bandwidth of this function is $\sim 0.9/N\Delta t$. The minimum bandwidth depends on the fading spectrum bandwidth, typically 0.1-2 Hz, which fixes $N\Delta t$, the maximum integration time.

The noise is uncorrelated over time intervals comparable with realiable pulse repetition periods (Δt) so quite large values of N may be used. The lower limit of Δt is determined by the presence of multiple reflections. During the day pulse repetition periods of ~ 10 ms can be used without ambiguities due to multiple reflections so significant pulse integration is possible. At night the acceptable repetition periods are much longer because of the large number of

multiple F-region reflections, so N must be smaller.

VINCENT and BALL (1981) found that the use of integration following a
coherent detector increased the signal-to-noise ratio in partial reflection
spaced antenna winds observations so that the minimum height for continuous
measurements over 24 hours was decreased from 82-84 km to 80 km. HOCKING's
(1981) measurements at Townsville (Figure 2) show ~9 dB improvement expected
when 8 pulses from a coherent detector are integrated.

DATA SYSTEM PARAMETERS

The data system takes the output from the radar system detectors and
converts it to digital values suitable for input to the analysis methods
discussed by Briggs in an accompanying article.

(a) Data Precision

If the data is to be used only for the full correlation analysis of fading
records from the spaced antennas, it is only necessary to provide one binary
digit (bit) from each receiver for each height. The resulting calculation is
variously called tetrachoric, polarity-coincidence, sign, or clipped signal
correlation. The resulting correlation functions are entirely satisfactory for
estimating middle atmosphere winds (FRASER, 1965; 1968; MEEK and SOIFERMAN,
1979).

If coherent detectors are used the one-bit signal is obtained by following
the detectors by comparators whose reference input is grounded. The data bit
then represents the sign of the detector output (FRASER, 1965). If incoherent
detectors are used, the one-bit signal may be obtained by following the
detectors with a comparator whose reference input is the mean detector output.
The data bit then indicates whether the detector output is above or below its
mean value. If a digital processor is available the one-bit data can be
obtained directly, after a/d conversion, by subtracting the mean value from the
raw data and retaining only the sign bit (MEEK and SOIFERMAN, 1975).

The tetrachoric correlation function is not in general simply related to
the correlation function of the unclipped waveform. If a waveform with a normal
probability distribution of amplitude with zero mean and a time autocorrelation
function $r_n(\tau)$, is clipped about its mean value of zero, then the tetrachoric
autocorrelation function, $r_c(\tau)$, is given by

$$r_c(\tau) = (2/\pi) \sin^{-1} r_n(\tau) .$$

MCFADDEN (1956) gave an approximate expression for the tetrachoric correlation
function of a sine wave added to the normally distributed waveform.

The position of cross-correlation maxima will not change when clipping is
used but some second-order changes in correlation parameters may result. As the
characteristics of the unclipped signal are not usually measured a transforma-
tion cannot be applied to deduce the correlation function of the unclipped
signal.

The advantage of 1-bit data is that the operation of multiplication is
replaced by AND or EXCLUSIVE or logical operations which are usually one-
instruction operations. However, the increasing speed of electronic circuits
has reduced multiplication time to values less than the instruction cycle time
of small computers and the use of tetrachoric correlation functions no longer
provides an advantage in computing speed.

(b) Dynamic Range and Analogue-to-Digital Converter Resolution

Even if 1-bit accuracy is sufficient for the determination of the correlation functions, the conversion of amplitude data from incoherent detectors still requires an estimate of the mean value. This can vary considerably from the weakest reflections close to the noise level to, possibly, the E-layer total reflection under conditions of low absorption.

At any site the receiver noise level is determined by sources of man-made noise, natural noise such as thunderstorm activity and by the thermal noise of the receiver input stages (including any ionospheric thermal radiation). On some occasions the minimum noise level will be determined by the thermal noise. The maximum receiver output will be determined by the receiver overload characteristics. The required dynamic range in the a-d conversion is therefore $S = V_{max}/V_n$ where V_{max} is the receiver output overload voltage and V_n is the mean noise level. If n-bit accuracy is required for the measurement of V_n then the required a-d resolution is

$$m = \log_2 (V_{max}/V_n) + n \text{ bits} .$$

The minimum level to which the noise can be set is determined by the detector characteristics. Typically, an active rectified detector has a lower limit of ∿3 mV and the full-scale value for an a-d converter is 10 V. The lowest feasible limit for n is 1 so m = 13. at least. Standard a-d converters only have 8 to 12-bit resolution so it is desirable to use receiver attenuators to reduce the number of bits required.

(c) Sample-and-Hold Requirements

Sample-and-hold (or track-and-hold) circuits are essential because the variation in signal amplitude over the a-d conversion time will be greater than 1 LSB. For a 10 bit a-d with a maximum input of 10 V, the voltage change corresponding to the LSB is 10×2^{-10} V ≃ 10 mV. If the conversion time is 10 μs, the maximum slope of the converter input signal is 10 mV in 10 μs = 1000 V sec^{-1}. If that signal is a Gaussian pulse with an $e^{-1/2}$ width of 15 μs (FWHP = 25 μs) and a peak amplitude of 10 V, then its maximum rate of change is ∿10^{6}V sec^{-1}, which exceeds the a-d conversion ?. Consequently, the signal must be held and the a-d conversion started after the track-and-hold output has settled.

(d) Logarithmic Post-Detector Amplifiers

The a-d conversion process introduces quantisation noise with an rms value of $\Delta V_1/ \sqrt{12}$, where ΔV_1 is the voltage change represented by the least significant bit. It is desirable that the weaker reflections should have the least amount of quantisation noise added, so some form of non-linearity is desirable, with greater amplification for the weaker reflections. If the quantisation noise is to represent a constant fractional error at all signal levels the a-d converter must be preceded by a post-detector amplifier whose output (V_o) changes are

$$dV_o \propto \frac{dV_s}{V_s}$$

where V_s = input signal level, i.e.,

$$V_o \propto \log V_s .$$

A logarithmic post-detector amplifier is also useful when checking transmitter power and height calibration. The transmitter output pulse is measured (by a receiver preceded by adequate attenuation) to ensure that its amplitude *and its position as the zero for height calibration* are within the required limits. Measuring the transmitter pulse through one of the receivers has the advantage of not requiring correction for the delay introduced by the receiver bandwidth. A parabola fitted to the a-d output data is equivalent to fitting a Gaussian function to the transmitter pulse, making a very good approximation: If

$$V_s = V_o \exp(-t^2/2t_p^2)$$

then

$$V_v \propto \log V_s, \quad \propto \log V_o - t^2/2t_p^2 \ .$$

There is no need to introduce an anti-logarithm function after the a-d conversion as the logarithmic conversion is a monotonic process and will not introduce any significant errors – assuming that the space and time correlation functions do have the same shape. A comparison (SMITH, 1981) of the winds measured with and without the anti-logarithm correction confirms that the error is negligible.

REAL-TIME ANALYSIS

There are substantial logistic problems in handling raw data from a multiple-height partal reflection spaced antenna wind system. A continuously running system will produce 1 to 3 1200 foot magnetic tapes in a day. Transfer of magnetic tapes to centralized computing facilities may be time consuming if large distances are involved and since a considerable amount of the data may be merely receiver noise or interference, it is wasteful of time and funds.

A dedicated small computer is considerably more efficient as only useful data is completely processed and the final output data is more compact. It also allows the possibility of interactive processing for development studies, the observation of correlograms on x-y displays, equipment calibration, and can still record raw data if required for special analyses. Software can also be used to change the hardware functions to measure complementary parameters such as partial reflection electron densities or ionospheric absorption.

(a) Real-Time Installations

Real-time partial reflection wind systems have the same basic requirements for the on-line computer(s), namely:

(1) control of transmitter pulsing;
(2) control of antenna switching;
(3) adjustment of receiver attenuation;
(4) initiation of the analogue-to-digital conversion of the required heights;
(5) calculation of the correlation functions;
(6) estimate of the wind velocity from the correlation function.

The basic requirements may be met in various ways. For example, the Christchurch installation has three parallel receiving channels (from antenna to a/d converter) whereas other installations switch one receiving channel around the receiving antennas; the Saskatoon and Urbana installations use two microcomputers in contrast with the single minicomputers used at Adelaide and Christchurch.

The Adelaide installation (VINCENT and BALL, 1981) uses coherent detectors followed by integration, and 8-bit data resolution. The computer is a Data General Nova.

At Saskatoon (MEEK and SOIFERMAN, 1979) incoherent detectors are used. The data resolution is 8 bits, followed by subtraction of mean values and conversion to 1-bit format. An APPLE II microcomputer is used for control, data collection and the calculation of the tetrachoric correlation functions which are transferred directly to a CBM PET microcomputer for the winds analysis.

At Urbana (RUGGERIO and BOWHILL, 1982) incoherent detectors are used, followed by integration, and the data resolution is 8 bits. Hardware control, data collection and partial calculation of the correlation functions are carried out by an APPLE II plus microcomputer with an arithmetic processor. A second APPLE II microcomputer completes the calculation of the correlation functions and carries out the wind analysis.

(b) An Example of Real-Time Correlation Analysis

The installation near Christchurch, at the Birdlings Flat field site, of an on-line computing system for transmitter and receiver control, data collection and analysis, is used principally for the full correlation analysis of partial reflection fading records (FRASER, 1973; SMITH, 1981). It is based on a DEC PDP-8/e minicomputer with paper tape and magnetic tape for input and output. Programming is carried out in a version of Algol-60 (ABBOTT, 1980) which allows ready access to assembler language subroutines and has a compact compiler output pseudocode. Subroutines for transmitter pulsing, for the control of attenuators and receivers and for the calculation of means and covariances are coded in assembler language. Computation speed was increased by the addition of a simple hardware add-and-shift multiplier/accumulator. The PDP-8/e will shortly be replaced by an LSI-11/23 and the software will be transferred, with new versions of assembler language subroutines.

Data is collected for 64 seconds with a sampling interval of 250 ms. The three logarithmic post-detector amplifier outputs are held by three track-and-hold circuits while the three 10-bit, 10 μs a-d converters convert the data to digital format. The a-d converter outputs are stored in three 10-bit latches and transferred, during the next a-d conversion cycle, through direct-memory-access into the computer. The triple hold-and-convert sequence is repeated at a rate of 6×10^4 sec^{-1}, giving a height resolution of 2.5 km (the transmitter pulse width is equivalent to a height increment of 4.5 km). The hold-and-convert sequence is controlled by the external height reference pulses to which the transmitter pulse is synchronized. A triple latch preceding the d.m.a. transfer copes with the jitter between the height reference pulses and the computer clock. Computer output pulses are, where necessary, synchronized to the height reference pulses with RS flip-flops.

The entire experiment is controlled by the computer and can recover automatically after a power failure. Provision is made for manual intervention, through the computer switch register and the console keyboard, to carry out tests such as continuous transmitter pulsing at 50 sec^{-1}, resetting of attenuators, dumping of data for the results of intermediate calculations, the final results of the correlation analysis and the display of the correlation functions through digital-to-analogue converters to an X-Y oscilloscope.

As the receiver dynamic range is limited to 60-70 dB, the height range from 65-102.5 km is divided into three intervals and the attenuator settings are changed by the program so that echoes within a given interval do not saturate the receivers.

The normal operating sequence is:

(1) Transmit a sequence of 256 pulses at 250 ms intervals, storing the 3 echo amplitudes for a succession of 6 heights at 2.5 km intervals.

                                                    Time: 64 seconds

(2) Calculate the mean values for one height sample and subtract them from the raw data, calculate the three autocorrelation functions and determine the mean autocorrelation function.

                                                    Time: 7 seconds

(3) Calculate the three cross-correlation functions.

                                                    Time: 15 seconds

(4) Carry out the full correlation analysis.

                                                    Time: 5 seconds

The total processing time is thus 27 seconds for each height sample. Data which do not satisfy the acceptance criteria are rejected as soon as possible. The minimum processing time, for a height sample containing only noise, is about 5 seconds. The minimum time to sample and process data for a group of 6 heights with no echoes is about 90 seconds. Even if signals are present, excessive interference or idiosyncratic diffraction patterns mean that not all heights yield an acceptable wind measurement. Thus the average processing time is 100-120 seconds for 6 heights. The observation rate could be increased by overlapping data collection with processing of the previous data but this requires extra data buffers and interrupt routines which have not been implemented.

(c) Languages for the Implementation of the Correlation Analysis

The above installations all use a mixture of low-level and high-level programming languages. Low level languages (e.g., assemblers or FORTH) are generally used for data collection but the computer program which carries out the correlation analysis will include a substantial amount of floating point arithmetic and function evaluation. It is highly desirable that a programme of this type should be written in a high-level language because

(1) Algorithms for floating point arithmetic, standard functions, I/O operations and error-checking are already available.

(2) The language is readily comprehensible to all associated with the project. This is particularly true in a university with the continual movement of students through a research project.

(3) The machine independence allows the application to the correlation analysis of known and well-tested algorithms used on other machines.

(4) Programs can be written and debugged more quickly.

(5) If the correlation analysis is being carried out in real time at a remote field site, program development can be carried out on any convenient machine and allow the field equipment to continue data collection (the same advantage is possible with assembler-language programs if a cross-assembler is available).

(6) Machine independence allows the replacement of part, or all, of the hardware without unnecessary re-programming in a new assembler language.

However, the significant advantage of assembler-language programming is the speed with which a specific algorithm can be implemented for specific hardware.

This is particularly true of vector operations like the accumulation of sums of cross-products for calculating the multiple correlation functions, and of responding quickly to external events.

The most effective language is therefore a high-level one which has straightford provision for calls to assembler-language subroutines and for transferring arguments to these subroutines. During programme development these assembler-language subroutines can be simulated in the high-level language.

CONCLUSIONS

PR spaced antenna radars have already made significant contributions to studies of the dynamics of the mesosphere and lower thermosphere (e.g., see VINCENT (1984) for a recent review). As an example of the height range covered, Figure 5 shows the mean zonal and meridional winds measured at Christchurch (44°S) and Scott Base (78°S). Although the Antarctic station is moderately small in terms of transmitter power and antenna size useful results are obtained down to heights near 75 km. Although high transmitter powers and large antennas are desirable for good system sensitivity, significant improve-emtns in this regard can be achieved at little cost by making the system phase coherent and coherently integrating the received signals. In this way relatively inexpensive PR systems can be constructed. VINCENT and BALL (1981) give an example of a transportable PR radar. The use of PR systems in the form of a network can be expected to lead to further advances in our understanding of the mesosphere.
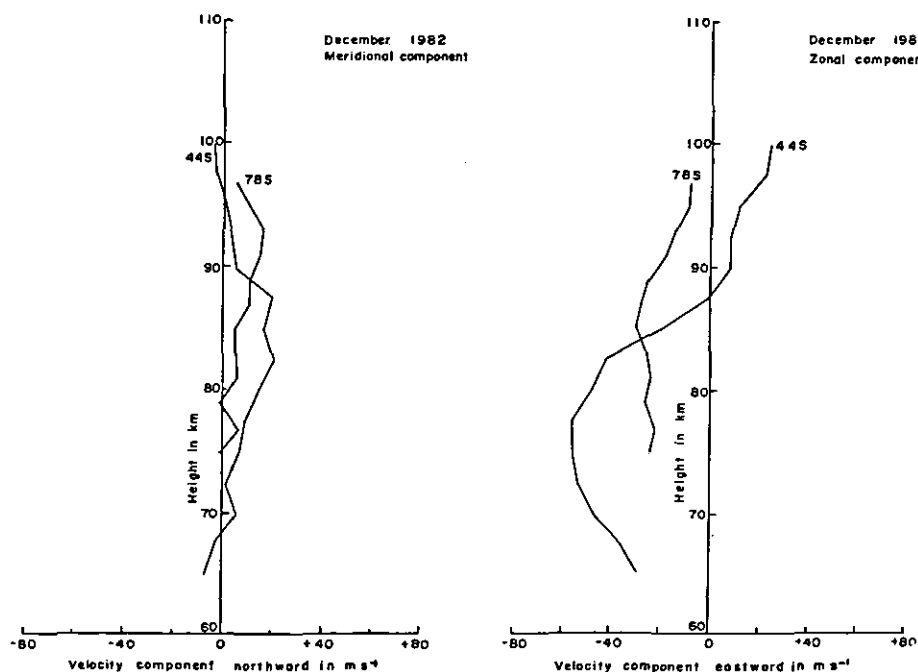

Figure 5. Mean meridional and zonal wind profiles at Christchurch and Scott Base.

REFERENCES

Abbott, R. H. (1980), The role of ALGOL 60 in computer education, Computer
    Educ., 35, 24.
Barber, N. F. (1957), Timing errors in an equilateral array of receivers, J.
    Atmos. Terr. Phys., 11, 299.
Belrose, J. A. (1970), Radio wave probing of the ionosphere by the partial
    reflection of radio waves (from heights below 100 km), J. Atmos. Terr.
    Phys., 32, 567.
Briggs, B. H. (1980), Radar observations of atmospheric winds and turbulence: a
    comparison of techniques, J. Atmos. Terr. Phys., 42, 823.
Budden, K. G. (1961), Radio Waves in the Ionosphere, Cambridge University
    Press, Cambridge.
Felgate, D. G. (1970), On the point source effect in the measurement of iono-
    spheric drifts, J. Atmos. Terr. Phys., 32, 241.
Felgate, D. G. and M. G. Golley (1971), Ionospheric irregularities and movements
    observed with a large aerial array, J. Atmos. Terr. Phys., 33, 1353.
Fraser, G. J. (1965), The measurement of atmospheric winds at altitudes of 64-
    120 km using ground-based radio equipment, J. Atmos. Sci., 22, 217.
Fraser, G. J. (1968), Seasonal variation of southern hemisphere mid-latitude
    winds at altitudes of 70-100 km, J. Atmos. Terr. Phys., 30, 707.
Fraser, G. J. (1983), On-line control and data collection for ionospheric
    experiments, 9th New Zealand National Electronics Conference, Christchurch,
    New Zealand.
Fraser, G. J. (1983), Summer circulation in the Antarctic middle atmosphere,
    J. Atmos. Terr. Phys. (in press).
Fraser, G. J. and R. A. Vincent (1970), A study of D-region irregularities,
    J. Atmos. Terr. Phys., 32, 1591.
Gardner, F. F. and J. L. Pawsey (1953), Study of the ionospheric D-region using
    partial reflections, J. Atmos. Terr. Phys., 3, 321.
Golley, M. G. and D. E. Rossiter (1970), Some tests of methods of analysis of
    ionospheric drift records using an array of 89 aerials, J. Atmos. Terr.
    Phys., 32, 1215.
Golley, M. G. and D. E. Rossiter (1971), Some aspects of ionospheric drifts
    using partial and total reflections from the lower ionosphere, J. Atmos.
    Terr. Phys., 33, 701.
Gregory, J. B. (1956), Ionospheric reflections from heights below the E region,
    Aust. J. Phys., 9, 324.
Gregory, J. B. (1961), Radar wave reflections from the mesosphere. 1. Heights
    of occurrence. J. Geophys. Res., 66, 429.
Gregory, J. B. (1963), Particle influx at high latitudes. 2. Solar protons,
    J. Geophys. Res., 68, 3097.
Gregory, J. B. and D. T. Rees (1971), Wind profiles to 100 km near 53°N during
    1969, J. Atmos. Sci., 28, 1079.
Gregory, J. B., C. E. Meek and A. H. Manson (1982), An assessment of winds data
    (60-110 km) obtained in real-time from a medium frequency radar using the
    radio wave drifts technique, J. Atmos. Terr. Phys., 44, 649.
Hocking, W. K. (1981), Investigations of the movement and structure of D-region
    ionospheric irregularities, Ph.D. Thesis, University of Adelaide, Adelaide,
    Australia.
Lawson, J. L. and G. E. Uhlenbeck (1950), Threshold Signals, McGraw-Hill Book
    Co. Inc., New York.
McFadden, J. A. (1956), The correlation function of a sine wave plus noise after
    extreme clippings, I.R.E. Trans. Inf. Th., IT-2, 82.
Meek, C. E. and J. Soiferman (1979), A real time ionospheric drift system,
    Institute of Space and Atmospheric Studies, Saskatoon, Canada.
Mitra, S. N. (1949), A radio method of measuring winds in the ionosphere,
    Proc. I.E.E., 96, 441.
Panter, P. F. (1965), Modulation, Noise and Spectral Analysis, McGraw-Hill
    Book, Co., New York.

Rottger, J. (1981), Investigations of lower and middle atmosphere dynamics with spaced antenna drifts radar, J. Atmos. Terr. Phys., 43, 277.

Ruggerio, R. L. and S. A. Bowhill (1982), New advances in the partial-reflection-drifts experiment, Aeronomy Laboratory, University of Illinois, Urbana.

Smith, M. J. (1981), Upper atmosphere circulation and wave motion, Ph.D. Thesis, University of Canterbury, Christchuch, New Zealand.

Starr, A. T. (1953), Radio and radar technique, Pitman, London.

Stubbs, T. J. (1973), The measurement of winds in the D-region of the ionosphere by the use of partially reflected radio waves, J. Atmos. Terr. Phys., 35, 909.

Van Vleck, J. H. and D. Middleton (1946), A theoretical comparison of the visual, aural and meteor reception of pulsed signals in the presence of noise, J. Appl. Phys., 17, 940.

Vincent, R. A. (1984), MF/HF radar measurements of the dynamics of the mesopause region – A review. J. Atmos. Terr. Phys. (in press).

Vincent, R. A. and S. M. Ball (1981), Mesospheric winds at low- and midlatitudes in the southern hemisphere, J. Geophys. Res., 86, 9159.

Vincent, R. A. and J. Rottger (1980), Spaced antenna VHF radar observations of tropospheric velocities and irregularities, Radio Sci., 15, 319.

Vincent, R. A., T. J. Stubbs, P. H. O. Pearson, K. H. Lloyd and C. H. Low (1977), A comparison of partial reflection drifts with winds determined by rocket techniques, J. Atmos. Terr. Phys., 39, 813.

16. D1 LF WIND MEASUREMENTS IN THE 90 TO 100 KM HEIGHT RANGE

R. Schminder and D. Kurschner

Karl Marx University Leipzig
Department of Physics, Geophysics Division
Collm Geophysical Observatory
DDR-7261 Collm, GDR

HISTORICAL NOTES

The first attempt to conduct D1 drift measurements using the explanation of
the phenomena which basically continues to apply to this very day, was made by
PAWSEY (1935). This work, however, was not carried forward at the time. The
measuring and evaluation method came to be known under the name of MITRA (1949)
because the experiments made by Krautkramer in Germany during World War II
remained secret and results were not published until 1950. The "similar fade"
method used for evaluation very soon showed a number of weaknesses, and these
were particularly striking in cases where the time displacements measured showed
a high degree of variability. When averaged with time shifts having different
signs, these cases sometimes resulted in very low averages to which a high velo-
city had to be assigned, whereas logical consideration suggested that the true
velocity was very low. With their definition of the "chaotic" or "character-
istic" velocity v_c BRIGGS et al. (1950) described the development of full
correlation analyses which may provide a complete description and space-time
analysis of the movement of fading amplitude patterns across the earth's
surface. Initially, however, technical problems arose in the implementation of
the method because the use of computers was limited, particularly in the case of
3^d analyses (GUSEV and MIRKOTAN, 1960). Other approaches which had been taken
in the meantime (e.g., PUTTER, 1955) proved impractical, while the discussion
which centered around the dispersion in the fading records (velocity increase
with the fading rate) found by JONES and MAUDE (1965) and given the one-sided
interpretation as wave phenomena, tended rather to hinder the spread of iono-
spheric drift measurements. The mid-60s finally brought the advent of computer
systems which enabled the widespread use of full temporal correlation analysis.
Methodical studies showed, however, that correlation analysis requires careful
consideration of a number of problems (e.g., high-pass filtering, triangle size
effect (CHANDRA and BRIGGS, 1978; CHANDRA, 1978)). At the same time, analyses
based on a modified and algorithmized type of similar fade evaluation (see
below) gave results with a high level of agreement and demonstrated a number of
technical advantages of the similar fade method which were not limited to its
easier implementation. The central problem of obtaining reliable absolute
values for the true velocity was also solved with the aid of similar fade
evaluation, while the parameters of the amplitude pattern which can be deter-
mined only through correlation analysis have met with little interest in the
past anyway. In medium latitudes the amplitude patterns are almost isotropic.
In addition, an emerging meteorology of the high atmosphere directed attention
increasingly to the height range under 100 km which, for the time being, was
accessible to the D1 method only via the partial reflection drifts technique.

What is needed in this height range which is of some relevance to
meteorology, is the regular and continuous patrol of wind fields and their
regular (seasonal) and irregular (disturbed) variations in a global synoptic
measuring network. "It is concluded that circulation changes in the lower
atmosphere significantly affect the upper atmosphere at least up to E-region
heights" (SCHWISTER and EBEL, 1981). Ground-based measurements will be given
first choice in this connection. Apart from the widely used radar meteor wind
observations (D2), the above-mentioned so-called partial reflection drifts
technique and the MST radar facilities which are being introduced on a larger

scale, drift (wind) measurements using total reflections with the aid of low frequency (LF) broadcasting stations have proved particularly useful here. As early as the 1950s, MILLMAN (1951, 1952), BOWHILL (1956), CILLIERS (1960), SALES and BOWHILL (1962), LEE and BOWHILL (1963), and BEYNON and GOODWIN (1967) reported sporadic measurements, some using their own LF and VLF transmitters. The first systematic and long-term measurements were begun in 1954 by SPRENGER (1958) at the Kuhlungsborn observatory for ionospheric research in the GDR, followed by the Collm observatory in 1956. This approach has been consistently pursued at the Collm observatory to the present day and involved such stages as the rehabilitation of the similar fade method by introducing a modified evaluation procedure (SCHMINDER, 1964), the algorithmization and automation of the entire measuring and evaluation process (KURSCHNER and SCHMINDER, 1980) and the development of a new method for measuring absolute reflection heights of the sky wave from CW transmitters at oblique incidence used also for wind measurements (KURSCHNER, 1981).

On the whole, LF wind measurements are little practiced on an international scale. As far as we know, the only other place where they are used is the Badary observatory (near Irkutsk/USSR). This may be due to a certain bias against using conventional broadcasting stations which might be regarded as inadequate aids by some. Expensive facilities will attract considerable attention, and this is also a question of public relations for the sponsor. Two opinions on this: "Simple instruments are needed for the upper atmospheric equivalent of weather stations. The big radars can be used either alone or in conjunction with other instrumentation for comprehensive local study of the details of the dynamical processes that control the properties of the upper atmosphere" (WALKER, 1979), and "I am impressed with the results of such instruments as the MST radars but I would like to draw your attention to inexpensive methods for collecting data, e.g., the horizontal wind drift measurements in the D-region. I am very surprised that we have only two stations in the world using the D1 method in the longer wave range. Maybe it would be very useful to organize such stations in other places including locations where we have MST radar. This would be good for calibration as well as a very valuable tool for experimental information on the dynamics of the upper atmosphere" (KAZIMIROVSKY, 1981b).

MEASURING AND EVALUATION PROCEDURE

This is based on the spaced receiver method. Transmitters used for the measurements are conventional amplitude modulated LF broadcasting stations with powers of the order of 500-1000 kW. In all cases the polarization of the transmitted electromagnetic waves is vertical. The three receivers for the reception of the sky wave reflected from the lower ionosphere (oblique incidence) are placed at the corners of a right isosceles triangle with sides oriented towards E and N, or in arrays. For frequencies of 150-300 kHz and transmitter distances from 150-500 km the triangle sides should be between 400 m and 200 m ($\lambda/5$, λ = wavelength), but not greater than $\lambda/3$. This gives a good compromise between the conflicting requirements for the greatest possible measuring accuracy (time lags not too small) and a high level of correlation.

A very wide fading spectrum can be expected as will be gathered from conventional fading records. The range of useful periods is limited on the one hand by periods T of several minutes (symbol N1) and related restrictions inherent in the measuring equipment, and on the other by T = 2 Δt where Δt is the maximum time shift between fading extrema. The following simple relations exist under conditions where a is the leg length of the measuring triangle and, therefore, a/2 the distance between the reflection points, Δt_z and Δt_m being the time lags between the fading extrema in the zonal and meridional directions, v the instantaneous velocity and β the wind direction measured from N to E:

$$v = \frac{a_2}{\Delta t_z^2 + \Delta t_m^2} \qquad \beta = \arc \tan \frac{\Delta t_z}{\Delta t_m}$$

where $v_z = v \sin \beta$ and $v_m = v \cos \beta$ are the zonal and the meridional components of the wind. Two approaches are possible for obtaining averages over a period of time suited to the particular problem, which may not be less than 10 mins. and need not be more than 30 mins. for the frequencies and fading rates involved here. Averaging v_z and v_m allowing for signs gives an average v_d which can be identified with the true velocity of the full temporal correlation analysis. On the other hand, averaging Δt_z and Δt_m gives the apparent velocity v' which is also familiar from correlation analysis. If one follows the argument of PORTNYAGIN et al. (1978) a value for v_c is also obtained from the equation

$$v_c = v_d \sqrt{v'/v_d - 1}$$

which may be proportional to the turbulence velocity in the height range below 100 km. "An associated quantity, the 'characteristic' velocity v_c which describes the intrinsic change of the moving pattern, has not yet been embodied into dynamical studies" (GREGORY et al., 1979). "Little practical use has been made of this parameter, and it is therefore interesting to note that it is comparable to the irregular wind which is supposedly due to waves of period $\tau \geq 5$ min." (MEEK et al., 1979) (see also Briggs, this volume).

On-line automatic equipment can be used to determine v_z, v_m, and v_c, with a digital display of values. These wind data are corrected for semidiurnal tides to give a daily mean value. Although the measurements normally take place only at night, such separation will succeed at medium latitudes even in summer. In winter all-day measurements are possible for part of the time, but the great difference in height between the daylight and night-time reflection levels prevents 24-hour harmonic analysis which might yield the diurnal tidal wind component. The latter, however, is small at medium and higher latitudes anyway, and its short vertical wavelength makes it difficult to interpret so that this need not be a serious defect. On the other hand, the usefulness of the method as a whole may be greatly restricted at lower latitudes with their considerable diurnal tidal winds (DEVARA et al., 1981). The prevailing wind cannot safely be determined using only wind values measured at noon (ROTTGER et al., 1983; GREGORY et al., 1981).

TECHNIQUE OF AUTOMATIC WIND MEASUREMENTS

Ferrite rod aerials are used for receiving the sky wave (reception of the magnetic field component). The aerial inductance together with a parallel capacitance form a resonant circuit which is tuned to the reception frequency. The aerials are positioned in such a way that the zero point of the horizontal polar diagram is in the direction of the transmitter. The ground wave, which is constant with regard to field strength and polarization, is not received. Residual electrical ground wave component is received by the ferrite aerial in this zero position, independent of the direction, can be exactly compensated by means of an electrical auxiliary aerial (vertical rod), since the signals at the output of the electrical and magnetic aerials are phase-shifted by 180° to each other (so-called zero cleared aerials). Orientation and compensation of the ferrite aerials for a complete elimination of the ground wave is only possible during the daytime when the absorption of the sky wave is very large.

The aerials are connected to crystal-controlled narrow-band superheterodyne receivers. A high selectivity is achieved by using gyromagnetic filters in the IF section. The bandwidth to 3 dB is of the order of magnitude of 100 Hz at an

IF of 200 kHz. Due to the high selectivity, the interfering sidebands of the amplitude-modulated signal are completely suppressed and reception is limited to the carrier-wave frequency. The IF is transformed to 1 kHz in the receiver by a second oscillator and then forms the output after selective amplification. Thus the amplitude of this 1 kHz output signal is proportional to the field strength at the point of reception.

For automatic wind measurements, the output signals of the measuring receiver are supplied to a central point of the electronic equipment. In this automatic equipment the 1 kHz signals are filtered and amplified using automatically controlled LF amplifiers (time constant of gain control $\geq$ 10 min.) and transferred to an optimal amplitude range. The gain control eliminates fluctuations with very long periods only, due to the regular slow nighttime variations of the mean absorption level. Shorter periodic fading components on the contrary are without change. Linear rectifiers transform the 1 kHz signals to variable DC signals which vary like the fading of the field strength at the corresponding receiving points. Short periodic disturbance components, resulting, for instance, from atmospherics or other influences are suppressed by active low pass filters (cut off frequency 0.1 Hz). Then the DC signal voltages, optimally prepared in this way for sensitive automatic measurements are electronically differentiated by means of active analogue differentiation devices to identify the fading extrema. This represents a process which can be interpreted as a special kind of high pass filtering.

The differentiation is followed by transformation of the differentiated signals into binary functions using comparators and discriminators (analogue-to-digital conversion). The conversion makes allowance for the sign of the derivative and a simple binary valuation (one digit) of its absolute value only. The binary signals are the input variables of a sequential logic circuit which implements a number of fixed programmed rules, realizing a special type of digital coherence test to detect corresponding fading minima and maxima$_{(i)}$ in the variations of the three signals and to determine time shift pairs $\Delta t_z^{(i)}$ and $\Delta t_m^{(i)}$ (z = zonal, m = meridional) in a digital manner with an accuracy of 0.25 s. Every individual time shift is supplied to a storage unit. A ROM connected with this unit assigns the individual velocity components $v_z^{(i)}$ and $v_m^{(i)}$ to each time shift pair (accuracy 5 m/s). All computed individual velocity components, and the time shifts, are summed up and stored in two storage devices up to the data output by a digital printer according to a given time program, under normal conditions every 30 minutes. The output consists of the mean time shifts and velocity components, the number of individual measurements and the time of the measuring interval. The simplified structure of the automatic wind recorder is shown in Figure 1. The path of information is marked by straight lines, and the control signals by dashed lines. Control unit 1 controls the measuring of individual values, control unit 2 releases the output of the final data.

ABSOLUTE REFLECTION HEIGHT MEASUREMENTS IN THE LF RANGE

(a) Method

Measuring the absolute reflection height when using conventional LF broadcasting stations at oblique incidence has been an unsolved problem until recently since no method is known to measure the travel time difference between sky wave and ground wave at the point of reception, from which the reflection height could be determined immediately. Relative phase observations between the carrier-wave frequencies of sky wave and ground wave, which can be carried out technically simply, do not solve the problem because the order of magnitude of the quotient of travel time differences and wave lengths in the LF range ($\lambda \leq$ 2 km) excludes a definite height allocation.
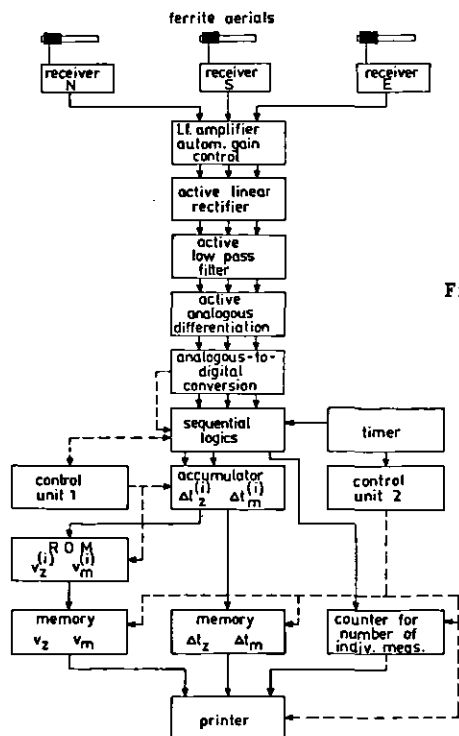
Figure 1. Block diagram of the automatic LF wind measuring equipment.

In 1981 Kurschner of the Collm Geophysical Observatory (GDR) was able to methodically elaborate and technically solve this problem. The essential principle of this new method is a transition from HF phase angle measurements to a time difference measurement between the LF modulation phases at selected phase points in a limited LF range. Since one individual LF oscillation with its period T_L comprises a number of carrier oscillations, it is possible to definitely determine reflection heights which require travel time differences up to a value of $\leq T_L$ with a given measuring path. If, on the other hand, a value is specified for a reasonable maximum height it is possible to determine an upper modulation frequency at which this height can just be determined definitely.

In the case of transmitter distances of 200 km and maximum heights of 150 km, this upper frequency limit is between 1 and 2 kHz (travel time differences of approximately 300 μs). Since portions of the frequency band selection for measurements are not permanently available in the modulation spectrum, height measurements are not completely continuous; however, frequently they are available in a rapid succession (measuring density 10^3-5.10^3 values per minute). The accuracy of travel time difference measurements required should be on average 10 μs, for a height resolution of 2 km. Related to a mean frequency of the modulation range of 1.7 kHz used, this approximately corresponds to one 70th of the period length. In order to achieve this measuring accuracy, improving the signal/noise ratio in the LF range is necessary by selecting a 200 Hz LF bandwidth. The exact determination of the parameters for this frequency range resulted from a study of the disturbing influences. These primarily include (a) the superposition of the ground wave component by sky wave components which cannot be excluded completely for technical reasons, and

(b) the superposition of ordinary and extraordinary magneto-ionic components especially under disturbed geomagnetic conditions with stronger trace splitting. When investigating the linear superposition of coherent amplitude-modulated signals which were time-shifted relative to each other, we could show that the errors due to (a) for a given range of reflection heights are negligible when selecting a range of modulation frequencies which can be determined theoretically. In addition, the measuring error, if any, due to (a) and (b) is completely eliminated, independent of the LF frequency, if the HF carriers of the superposed components assume relative phase values of approximately 90° and 270°. Between these points, the error passes through a positive and negative maximum value depending on the height and the LF frequency. Positive and negative error components are compensated by fluctuations of the carrier phases during a sufficient period of averaging (30 min), and the reflection height of the dominating magneto-ionic component can be determined with sufficient accuracy.

(b) Implementation of the Measuring Principle

    The system for height measurements consists of the following main units: (1) a wide-band ferrite aerial for receiving the sky wave (refer to earlier section) and a vertical rod aerial for receiving the ground wave; (2) a two-channel receiver, band width 4 kHz, including demodulators for supplying the modulation signals; (3) a two-channel LF filter unit for selecting the frequency range chosen; (4) an automatic unit for measuring the travel time differences including a control device for implementing the given algorithm in order to digitally determine the time differences between the modulation envelopes, and, (5) a functional sub-assembly for allocating the travel time differences and reflection heights as well as a data storage unit coupled with a digital output of the mean values. Calibration of the system is made during the day-time when the sky wave is absent. The sky wave aerial is turned in such a way that it receives portions of the ground wave. Under these conditions the travel time differences must be zero. The method described here in a simplified manner has proven very successful since the completion of the measuring system in April 1983. It is a valuable and necessary method for completing the high-atmosphere wnd measurements, opens up new approaches for studying vertical wind field gradients in the upper mesopause region and in the lower thermosphere, and should also be of interest for other upper atmosphere measurements such as A3 absorption measurements.

ADVANTAGES AND DISADVANTAGES OF D1 LF WIND MEASUREMENTS

(1) Daily measurements are possible over long periods using comparatively simple equipment (even in comparison with D2). Very good information is obtained at low cost. The procedure is therefore suitable for network measurements as part of a system for the permanent study of regular and irregular variations of the wind fields at the upper boundary of the height range (between 90 and 100 km) which is of some relevance to meteorology. A condition is the existence of LF broadcasting stations (150-300 kHz) spaced 150-500 km apart.

(2) Measurements can be conducted on several measuring paths from one point, giving representative data on the average wind conditions over larger areas, e.g., Central Europe. The transmitters are broadcasting stations which can be used free of charge for years or decades and normally do not change their location or frequency. This eliminates the need to provide a transmitter and avoids causing interference.

(3) The zonal and meridional wind components can be measured simultaneously and in the same volume in the upper atmosphere. The high measuring density (1 value/min./measuring path or more) makes it possible to study wind variations with very short periods (generated by gravity waves) and to calculate turbulence

parameter. "Turbulence is important on vertical transport" (LINDZEN, 1981).

(4) There are reliable and accurate measurements of the reflection height (reference height for the wind observations). The reflection heights of the long waves show marked nighttime variation, in winter for instance with values arund 90 km early at night and late in the morning and maximum heights around 95 km after midnight. On individual days the medium nightly reflection level may vary between 90 and 100 km or more. There is correlation with geomagnetic activity. In the range from ΣK_1 (15-09 UT) = 10 to ΣK_1 = 30 heights increase by half a kilometer for each unit that is added to the sum of K_1. Heights above 110 km (with the extraordinary magneto-ionic wave component) are often reached in geomagnetic storms with little influence on the wind field itself. This makes it possible to derive vertical gradients of the horizontal wind field parameters over a wide height range.

As is in the case with most other measuring techniques, no allowance is made for the vertical wind velocity. "Measurements of mean vertical speeds, which range from cm per second in the troposphere to m per second in the lower thermosphere, are not yet practicable. The two related techniques, Doppler and spaced-antenna radar, are in principle capable of doing so, but there are economic as well as technical limitations" (GREGORY, 1981). Attempts made by DROBSCHEW (1968) to use D1 frequency diversity methods were not carried forward at the time. It would seem, however, that at least the direction of movement could be determined in this manner.

(5) For heights below 100 km there is no disagreement with the view that the electron drifts measured correspond to the neutral gas wind, both in their direction and velocity. There is no disputing the possibility of separating the prevailing wind, tidal motions and gravity waves from the values measured. Closely spaced receiver facilities at distances where the reflection points form a triangle with sides of 20-30 km length, might lead to the determination also of the phase velocity and propagation direction of the gravity waves.

(6) The complete arrangement (3 receivers, automatic measuring and evaluation apparatus, height measuring equipment) is suitable for mobile use. This would enable the urgently needed standardization of possible measuring networks, plus good compatibility with results from different methods such as radar meteor wind measurements and results from the partial reflection drifts technique. "Different methods of reducing spaced fading records have been followed by different workers and in the absence of any standard recommended method inter-comparison of results at different locations may become difficult" (CHANDRA, 1979). "Comparisons of winds are desirable, made by simultaneous observations in a common value" (GREGORY et al., 1981). Synoptic maps for users of the data require reliable absolute values for the velocity at a measured height.

(7) The records are available on-line. Automatically operated equipment allows the parameters of circulation to be provided with a time delay that is determined and limited only by the period of tidal winds and which is acceptable for a synoptic meteorology of the upper atmosphere.

These measurements would seem to have a future if the results can be made to fit the requirements of practical meteorology. As far as the prevailing wind is concerned, this called for the modelling of wind systems over the entire MST range. With regard to the tidal winds, useful connections must be found with the thermal regime of the middle atmosphere. "The participants in these programs are all observationalists, and more theoretical input is needed in these programs" (GELLER, 1981). Finally, the role of the gravity waves and planetary waves for the transport processes of energy in the meridional and vertical directions across the middle atmosphere must be understood. This

should then lead to a concentration of the organization and, eventually, the implementation of such work in the form of routine programs with the meteorological services of individual countries and with the WMO.

RESULTS OF Dl LF WIND MEASUREMENTS OVER CENTRAL EUROPE AT THE COLLM GEOPHYSICAL OBSERVATORY

(a) Frequencies, Measuring Equipment, Timing

At the present time Dl LF wind measurements take place on 179 kHz (transmitter distance 170 km), 227 kHz (transmitter distance 465 km) and 272 kHz (transmitter distance 400 km). The length of sides of the measuring triangles is 300 m, directed toward N and E; there is a 7-point array. The receiver output voltages are processed at a central point in fully automatic fashion, with absolute reflection height being simultaneously measured on 179 kHz. In order to obtain a representative picture of the average wind conditions over Central Europe for a height range of ±2 km around the middle reference height indicated, the results for all three measuring paths are summarized. In 1982 the integral measuring density was 3 values/min. During a time of 4,777 hours (= 55% of the hours of a year) a total of almost 10^6 readings were obtained and interpreted.

The measurements are limited to the time from shortly before sunset to shortly after sunrise. Although all-day measurements are possible in winter, harmonic analysis is not useful over the entire 24-hour period because of reflection height variations. On the other hand, the daylight time interval of 6 hours in winter is too short for separate harmonic analysis. The wind field parameters measured can be printed out individually and in various time steps up to 30 minutes. Equi-distant intervals of 30 minutes are normally preferred because this meets the need both for sufficient averaging and for reliable determination of the tidal wind parameters. The prevailing wind and tidal components are separated with the aid of a harmonic analysis program using a minicomputer. Fading records may be kept in the same way if necessary, as a way of reviewing the fading spectrum and keeping a check on the measuring system. Note - As far as we know, the only other place where Dl LF wind measurements are made is Badary/USSR. The frequency there is 200 kHz, the transmitter distance 140 km and a 500 m measuring triangle is used (KAZIMIROWSKY, 1981a).

(b) Examples of Results

In the following, a number of examples will be given to illustrate the potential of the measuring procedure and the viability of the results. Figure 2 is a typical example of undisturbed mid-winter wind conditions. The measuring period for which a sufficiently constant reflection height of 90 km was measured, was 16 hours. The prevailing wind was directed toward ENE, and the two semi-diurnal tidal wind components are practically equal. The phase lag between the zonal and meridional tidal wind components was 0.2 h greater than the theoretical (single mode) value of 3 h. The value normally obtained (monthly average) is somewhat lower. The integral measuring density for this particular night was 3.4 values/min. The short period velocity variations can be attributed to gravity waves. The reflection point triangle which results from the measuring paths on 179, 227, and 272 kHz is, however, too large (200 km) for an accurate determination of the phase velocity and propagation direction.

In cases where the day-to-day wind variations are sufficiently small over a longer period (another interesting phenomenon is the day-to-day variability of the semi-diurnal tidal wind parameters), an average daily variation can be calculated as is shown in Figure 3 for December 1982. Since the reflection heights for December 1982 were between 85 and 110 km, it is expedient to consider at
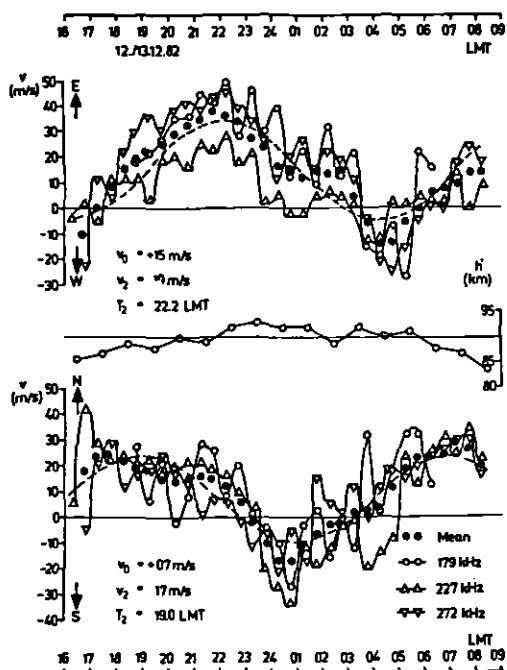
Figure 2. Night-time variations of the wind in the
upper mesopause region over Central Europe (52°N,
15°E) for an individual night in December 1982
obtained from LF wind measurements on 3 measuring
paths (v_0 = prevailing wind, v_2 = amplitude of
the semi-diurnal tidal wind, T_2 = phase of the
semi-diurnal tidal wind, defined as the mean local
time of the occurrence of the eastward and northward
wind maximum respectively).

least two groups of heights. If 94.5 km is taken to be the dividing line, then
the average values for the two groups of heights are 90 and 100 km,
respectively. The effects of the vertical gradients of the wind field para-
meters can be clearly seen – the phase of the semi-diurnal tidal wind for a
height of 100 km shows a shift of two hours toward the early evening compared
with a height of 90 km, and the zonal prevailing wind decreases with increasing
height as can be expected. If an average reflection height is calculated for
each night the vertical gradients can be determined from regression lines.
Figure 4 shows the conditions for the phase of the semi-diurnal tidalwind on all
three measuring paths (zonal only), and the average of all measuring paths
(zonal and meridional). Heights are measured on 179 kHz. For this measuring
path the vertical gradient of the phase is quarter of an hour per kilometer
(= 7.5°/km, 50 km vertical wavelength). The lower values for the other
measuring paths are due to the fact that the reflection height variations are
smaller for lower equivalent frequencies and for a steeper N(h') profile.

The prevailing wind and tidal motions show seasonal variations, and Figure
5 illustrates the annual variation of the wind field parameters measured. v is
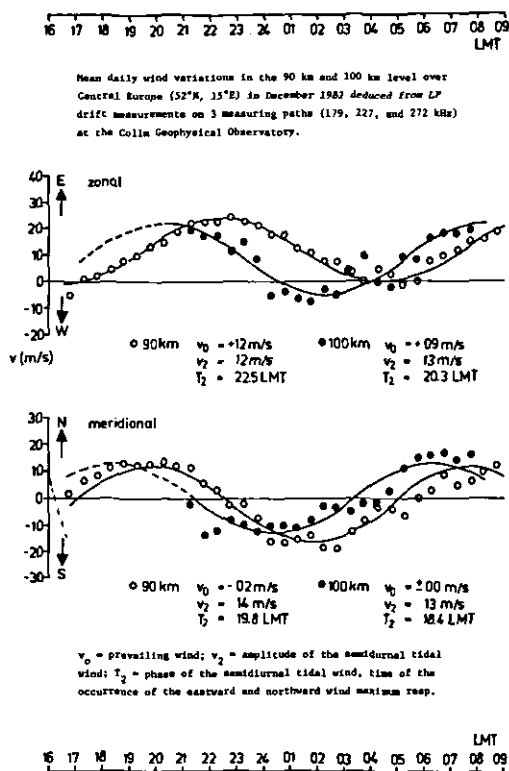the average of the moduli of the wind vectors. Most striking is the prevailing

Figure 3. Mean night-time wind variations in December 1982 for two different heights over Central Europe.
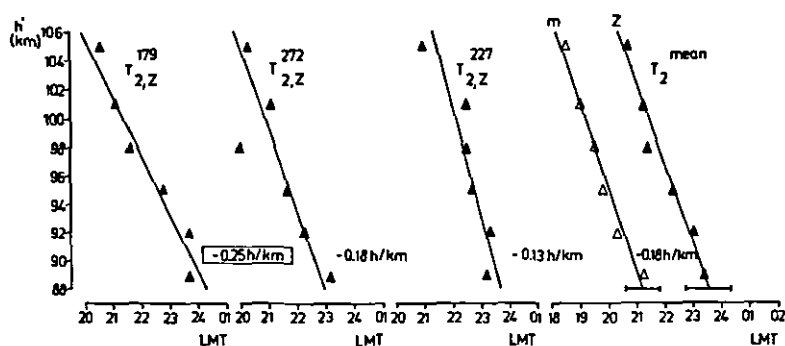

Figure 4. The vertical gradient of the phase of the semi-diurnal tidal wind in the 90 to 100 km height range over Central Europe in December 1982 and January 1983.
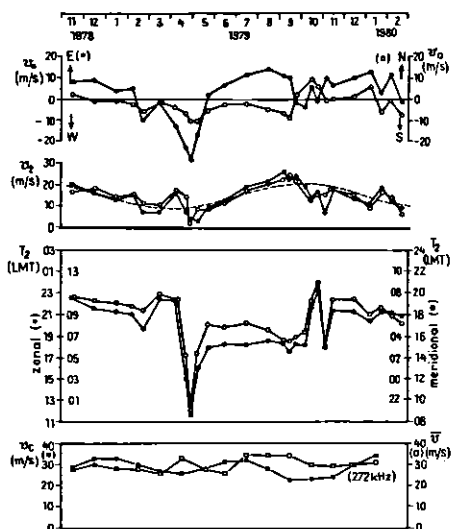
Figure 5. Seasonal variations of the wind field parameters in the upper mesopause region over Central Europe (taken from SCHmINDER and KURSCHNER, 1981).

wind in April which is directed toward WSW, as well as the abrupt phase changes in the semidiurnal tidal wind in spring and autumn. The annual variation of the amplitude of the semi-diurnal tidal wind shows minima in April and maxima in September, contrary to the annual variation of the ozone concentration. The September maximum is characterized by excellent reproducibility for each year (SCHMINDER and KURSCHNER, 1982) and an almost constant phase. The phase variations in spring and autumn mostly reach 2 (rotation or forward and reverse movement) but change from year to year and are not correlated in autumn and spring (SCHMINDER and KURSCHNER, 1979a).

In Figure 5 the zonal prevailing wind was directed toward W in February 1979. This was not in keeping with the season and practically anticipated the springtime circulation reversal as a result of the stratospheric warming effect in the upper mesopause region. The main feature of such a wind field disturbance is a reduction of the zonal, eastward directed prevailing wind and even a reversal of the wind direction resulting in a prevailing westward wind. In the various cases this prevailing wind reversal is accompanied by a shift in the phase and/or increased amplitude of the semi-diurnal tidal wind. The latter may also occur alone, i.e., without a perceptible effect on the prevailing wind if the primary disturbance of the middle atmosphere conditions does not reach the 90 km level (as happened in the winter of 1976/77; see SCHMINDER and KURSCHNER, 1979b). In some cases the southward directed meridional prevailing wind may increase. Figure 6 shows the conditions for the "stratwarm" event of February 1983. Over the period from February 8-17 the zonal prevailing wind changes from +07 m/s to -07 m/s (from west to east wind). There is a shift of 3 h in the phase of the semi-diurnal tidal wind, and the sign of the vertical phase gradient changes. The average nightly reference height was 93 km during both normal and disturbed periods. The v_2 values which appear in the figure show that. contrary to what has been said above, the amplitude of the semi-diurnal tidal wind has decreased. This is due to the averaging of the daily variations for tidal phases which have a comparatively large amount of variability.

Similar interesting phenomena in summer include the quasi- 2-day oscillation in the meridional and zonal prevailing wind. It was possible for the second time in 1982 to analyze these fluctuations and, this time, to observe
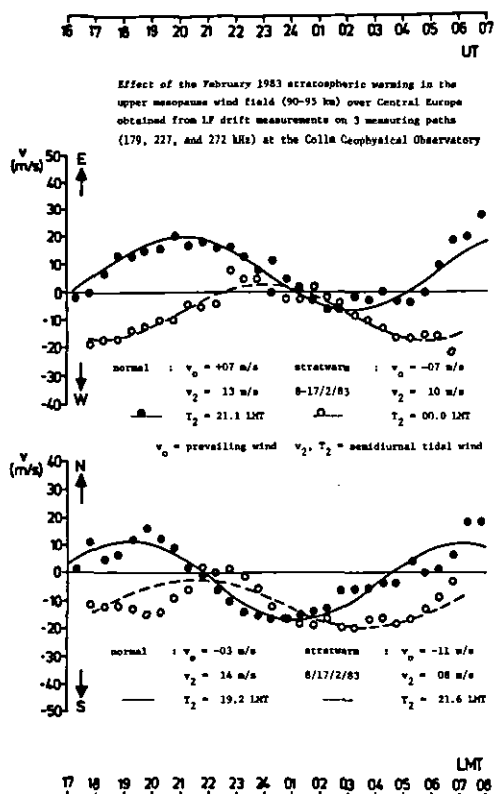
Figure 6. The wind field change in the upper mesopause region over Central Europe in connection with the 1983 February stratwarm effect.

them continuously from the end of May to the middle of August. Interest is rising internationally in this phenomenon, and more than a dozen papers have dealt with it in the last two years alone (e.g., CRAIG et al., 1980.

These few examples were designed to show that the scope and quality of the measurements meet the criteria which were established in the introduction as desirable and which have actually been implemented. In all cases where LF broadcasting stations are available in a suitable range of frequencies and distances the technique can be described as equal to other established measuring methods. The list of references has been kept to a minimum in view of the fact that a booklet, Schminder, R., "Bibliography of motions in the upper atmosphere" is to appear in 1984 published by Geod. Geophys. Veroff. NKGG of the GDR Academy of Sciences. Series 2. Berlin, with a list of about 3,600 titles, 900 of them from the Cyrillic literature.

REFERENCES

Beynon, W. J. G. and G. L. Goodwin (1967), Horizontal movements in the lower E- and D-region, J. Atmos. Terr. Phys., 29, 1181-1192.
Bowhill, S. A. (1956), The fading of radio waves of frequencies between 16 and 2400 kHz. J. Atmos. Terr. Phys., 8, 129-145.
Briggs, B. H. (1968), On the analysis of moving patterns in geophysics -I, J. Atmos. Terr. Phys., 30, 1777-1788.
Briggs, B. H., G. J. Phillips and D. H. Shinn (1950), The analysis of observations on spaced receivers of the fading of radio signals, Proc. Phys. Soc.,

B63, 106-121.

Chandra, H. (1978), On the triangle size effect in spaced receiver drift experiments, Indian J. Radio Space Phys., 7, 13-15.

Chandra, H. (1979), On the interpretation of velocities derived from spaced fading records, Proc. Indian Acad. Sci., A88, 41-49.

Chandra, H. and B. H. Briggs (1978), The effect of filtering on ionospheric drift parameters determined by full correlation analysis, J. Atmos. Terr. Phys., 40, 541-548.

Cilliers, W. A. (1960), The fading of low frequency radio waves reflected from the ionosphere. J. Atmos. Terr. Phys., 19, 102-114.

Craig, R. L., R. A. Vincent. G. J. Fraser and M. J. Smith (1980), The quasi 2-day wave in the Southern Hemisphere, Nature, 287, 319-320.

Devara, P. C. S., M. I. Ahmed, M. S. Rao and B. R. Rao (1981), Neutral wind measurements over Waltair using meteor radar, J. Atmos. Terr. Phys., 43, 239-242.

Drobschev, V. I. (1968), Investigations of the irregular structure of the iono-sphere using a combination of closely-spaced receiver method and frequency-diversity equipment (in Russ.), Ionos. Issl., 15, 141-146, Izd. Nauka, Moskva.

Geller, M. A. (1981), Tides, gravity waves, and turbulence, Handbook for MAP, Vol. 3, 44-50, edited by C. F. Sechrist.

Gregory, J. B., Middle atmosphere dynamics. Nature, 294, 519-520.

Gregory, J. B., C. E. Meek and A. H. Manson (1981), Mean zonal and meridional wind profiles for the mesosphere and lower thermosphere at 52°N, L = 4.4, during solar maximum, Atmosphere - Ocean, 19, 24-34.

Gregory, J. B., C. E. Meek, A. H. Manson and D. G. Stephenson (1979), Develop-ments in the radiowave drifts technique for measurement of high-altitude winds, J. Appl. Meteor., 18, 682-691.

Gusev, V. D. and S. F. Mirkotan (1960), 3^d full correlation analysis (in Russ.), Issl. Neodnorod. Ionos., 4, 7-19, Izd. Akad. Nauk, Moskva.

Jones, D. and A. D. Muade (1965), Evidence for wave motions in the E region in the ionosphere, Nature, 206, 177-179.

Kazimirovsky, E. S. (1981a), Dynamical regime of the lower thermosphere over East Siberia, Handbook for MAP, Vol. 2, 117-124, edited by S. K. Avery.

Kazimirovsky, E. S. (1981b), Handbook for MAP, Vol. 2, 496, edited by S. K. Avery.

Krautkramer, J. (1950), Uber wanderungserscheinungen rascher feld-starkeschwankungen von ionospharenechos archiv elektr., Ubertragung, 4, 133-138.

Kurschner, D. (1981), Methodical aspects and new tests for determining the reflection height of sky waves in the long wave range at oblique incidence using amplitude-modulated long-wave transmitters, Gerlands Beitr. Geophys., 90, 285, 294.

Kurschner, D. and R. Schminder (1980), Fortschritte bei der algorithmierung und strandardisierung der automatischen auswertung von ionospharendriftmessungen im langwellen-bereich und ihre bedeutung fur den aufbau von Me netzen zur synoptischen analyse hochatmospharischer windfelder geologie und geophysik, Geophys. Veroff., Karl Marx Univ., Leipzig, Akad. Verlag Berlin, Serie 3, Band 2, Heft 2, 219-227.

Lee, H. S. and S. A. Bowhill (1963), A statistical study of apparent horizontal ionospheric movements using 300 kc/s radio wave, J. Atmos. Terr. Phys., 25, 659-685.

Lindzen, R. S. (1981), Turbulence and stress owing to gravity wave and tidal breakdown, J. Geophys. Res., 86, 9707-9714.

Meek, C. E., A. H. Manson and J. B. Gregory (1979), Internal consistency analyses for partial and total reflection drifts data, J. Atmos. Terr. Phys., 41, 251-258.

Millman, G. H. (1951), A note on ionospheric wind measurements at 150 kHz, Ann. Geophys., 7, 272-274.

Millman, G. H. (1952), A study of ionospheric winds and turbulence utilizing

long radio waves, Ann. Geophys., 8, 365-384.

Mitra, S. N. (1949), A radio method of measuring winds in the ionosphere, Proc. Instn. Electr. Engrs., 96/III, 441-446.

Pawsey, J. L. (1935), Further investigations of the amplitude variations of downcoming wireless waves, Proc. Camb. Phil. Soc., 31, 125-144.

Portnyagin, Yu.I., K. Sprenger, I. A. Lysenko, R. Schminder, A. D. Orlyansky, K. M. Greisiger, Yu.D. Ilyicev, D. Kurschner and B. Schaning (1978), Ground-based wind measurements in the 90 to 100 km height range (in Russ.), Gidrometeoizdat Leningrad, 343 pp.

Putter, P. S. (1955), Messung des ionospharenwindes aus der wanderungs-geschwindigkeit eines zustandes (z.B. Echo-Feldstrake) langs der Erdoberflache, Rep. Phys. Soc., Conf. The Physics of the Ionosphere, Canvendish Lab. Cambridge, 191-201.

Rottger, J., P. Czechowsky, R. Ruster and G. Schmidt (1983), VHF radar observa-tions of wind velocities at the Arecibo observatory, J. Geophys., 52, 34-39.

Sales, G. S. and S. A. Bowhill (1962), Apparent ionospheric motion in the night-time D-region, J. Atmos. Terr. Phys., 24, 451-465.

Schminder, R. (1964), Ionospharische Driftmessungen im E-Schict niveau am Geophysikalischen Observatorium Collm - ein Beitrag zur Zirkulation der Hochatmosphare, Veroff. Geophys. Inst. Karl Marx Univ., Leipzig, Akad. Verlag Berlin, 2. Serie, Band 18, Heft 2, 79-275.

Schminder, R. and D. Kurschner (1979b), On the behaviour of wind systems in the upper mesopause region in winter and during the transition from winter to summer conditions, J. Atmos. Terr. Phys., 41, 119-122.

Schminder, R. and D. Kurschner (1981), Seasonal variations in the wind field of the upper mesopause region in 1979, Gerlands Beitr. Geophys., 90, 22-32.

Schminder, R. and D. Kurschner (1982), The seasonal maximum in the amplitude of the semi-diurnal tidal wind in the upper mesopause region over Central Europe, J. Atmos. Terr. Phys., 44, 207-210.

Schwister, B. and A. Ebel (1981), Propagation of circulation perturbations from the stratosphere into the lower ionosphere, Kleinheubacher Berichte, 24, 237-243.

Sprenger, K. (1958), Ionospharendriftmessungen im Langwellenbereich als Beitrag zum Problem der allgemeinen Zirkulation der Hochatmosphare, Z. Meteor., 12, 211-218.

Walker, J. C. G. (1979), Radar measurement of the upper atmosphere, Science, 206, 180-189.

CUMULATIVE LISTING FOR THE MAP HANDBOOK

Volume	Contents	Date of Publication
1	National Plans, PMP-1 Report; PMP-2 Report, PMP-3 Report, MSG-4 Report, Approved MAP Projects	*June 1981
2	Symposium on Middle Atmosphere Dynamics and Transport (extended abstracts)	June 1981
3	PMP-5 Report, MSG-1 Report, MSG-2 Report, MSG-3 Report, Antarctic Middle Atmosphere Project (AMA), EXOS-C Scientific Observations, WMO Report No. 5, Updated Chapter 2 of MAP Planning Document, Condensed Minutes of MAPSC Meetings	November 1981
4	Proceedings of MAP Assembly held in Edinburgh, 14-15 August 1981, Condensed Minutes of MAP Steering Committee Meetings held in Edinburgh, Proceedings of MAP Open Meeting held in Hamburg, 19 August 1981	April 1982
5	A Catalogue of Dynamic Parameters Describing the Variability of the Middle Stratosphere during the Northern Winters	May 1982
6	MAP Directory	November 1982
7	Acronyms, Condensed Minutes of MAP Steering Committee Meetings, Ottawa, May 1982, MAP Project Reports, National Reports, Committee Reports, PMP and MSG Reports, Workshop Reports, Announcements and Corrigendum	December 1982
8	MAP Project Reports: DYNAMICS, GLOBUS, and SSIM, MSG-7 Report, National Reports: Czechoslovakia, USA	July 1983
9	Papers presented at the URSI/SCOSTEP Workshop on Technical Aspects of MST Radar, May 23-27, 1983, Urbana	December 1983
10	Papers presented at the International Symposium on Ground-Based Studies of the Middle Atmosphere, May 9-13, 1983, Schwerin, German Democratic Republic	May 1984
11	Condensed Minutes of the MAP Steering Committee Meetings held in Hamburg 13-14 August 1983, Research Recommendations for Increased US Participation in the Middle Atmosphere Program, GRATMAP Project Report, MAP Study Group MSG-7 Report	June 1984
12	Coordinated Study of the Behavior of the Middle Atmosphere in Winter (PMP-1) Workshops	July 1984
13	Ground-Based Techniques	November 1984

GROUND-BASED TECHNIQUES

CONTENTS