

7.7A PROBLEMS AND SOLUTIONS IN ANALYZING PARTIAL-REFLECTION DRIFT DATA  
BY CORRELATION TECHNIQUES

C. E. Meek

Institute of Space and Atmospheric Studies  
University of Saskatchewan  
Saskatoon, Canada

## INTRODUCTION

The problem of analyzing spaced antenna drift data breaks down into the general categories of raw data collection and storage, correlation calculation, interpretation of correlations, location of time lags for peak correlation, and velocity calculation.

## DATA COLLECTION AND STORAGE

The total record length required for wind analysis depends on data stationarity, and may be different at different sites (or heights, or times). Generally periods of 2-5 min are used at medium frequencies (MF) and somewhat shorter records at VHF, depending on local conditions and equipment. In real-time systems, another consideration is whether analysis can be done during data collection or not; the former is preferable since longer data sequences lead to better stability in the final wind value. Since gravity waves of short period (5-10 min) are likely to have short horizontal wavelengths, these are probably attenuated by the wide antenna beam widths employed at MF, and so the longer records are reasonable.

Raw data storage is a problem, particularly if there are a lot of height gates. It is best to have no gaps in the time sequence for the whole record, so that later correlations can use the full length, and stability will not decrease appreciably with lag. Alternately, the gaps between data blocks can be set to be an integer number of  $\Delta t$  (the basic sample or cycle rate), and the data interpolated later. In real-time systems the data must be broken into blocks, which are dealt with as they are collected, by accumulating partial correlations for all the lags required.

If the transmitter pulse is generated from the local power line frequency, a phase shift can move it away from locally generated interference spikes. Otherwise some sort of despiking is required on the raw data. One easy method is to replace any amplitude which is more than the sum of the adjacent amplitudes by the average of these values, a simple process in microcomputers. If despiking is not done, and the spikes are well above the signal amplitude, then spurious peaks can occur in the cross correlations.

The sample rate depends on the expected velocities. A reasonable estimate is  $\Delta t = d/(2V_{\max})$ , where  $d$  is the antenna separation and  $V_{\max}$  is the largest expected velocity. In this case the cross correlation peak furthest from zero lag would be located at approximately 1/2 lag. Data with wide cross correlations would require faster sample rates, since the peaks are less well defined in lag.

A comment should be added here on receiver gain control. The dynamic range of signals (MF) is usually -60 dB over the mesosphere/themosphere heights. A voltage-controlled gain is usable, so that the gain can be changed with height, but if this is done in the IF stage the output will take at least one height gate to settle down (assuming that the height gates are chosen for maximum height resolution). A better way is to have a receiver with distributed outputs, each for a fixed attenuation, and select the appropriate output

for a particular gate by the signal strength. It is important to have some estimate of signal strength at each height gate in order that spurious interference, particularly teletype signals at night, can be recognized in examination of the final wind values. Such interference will be constant in all height gates, and can be rejected by choosing a height gate where no signal is expected, and setting a signal strength acceptance limit at some level above this.

#### CORRELATION CALCULATION

The normal definition of correlation is used with raw amplitudes, but data which has been converted to single bit amplitudes with respect to the mean must be corrected according to  $\rho_1 = \sin(\rho\pi/2)$ , where  $\rho$  is the correlation value of the single bit data.

#### INTERPRETATION OF CORRELATIONS

Examination of the auto- and cross correlations (ACF, CCF) is invaluable, particularly at a new site or if equipment is modified, for determining data stationarity, noise problems, interference pattern effects, etc.

The basic parameters required from the correlation are the time lags for maximum CCF  $t_{\max}$ , used for "apparent velocity", and peak values of the CCFs,  $\rho_{\max}$ , along with the width (in lag) of the ACF,  $t_A$ , used for the "true velocity".

In theory, the sum of the  $t_{\max}$ , taken in the proper sense around the receiver antenna pair vectors, should add to zero. In practice there are often multiple peaks, and even if there is only one but it is very wide, statistical fluctuations can make the accurate determination of  $t_{\max}$  difficult.

There are many possible causes for multiple peaks. If the record length is too short, trends in the signal strength can cause a high correlation value at zero lag, in addition to any real peaks. A wind shear over the antenna beam width or pulse width, or any abrupt change in velocity during the record will produce a separate set of peaks for each velocity present. If these changes were smooth rather than abrupt, the result would be an average correlation function with a single peak. Other causes of multiple peaks include fringe patterns caused by interference between two or three strong reflectors.

For systems using coherent averaging, it is possible that with the high pulse rate employed, multiple reflections or scattering from, say, the F layer, could contribute "spurious" peaks to the correlations. This problem can be alleviated by making each pulse independent in phase. In any case, a method is required by which only those peaks corresponding to one pattern motion be chosen for further analysis. A simple way is to use the normalized time discrepancy ( $NTD = |\sum t_{\max}| / \sum |t_{\max}|$ ) in the selection of peaks. A high value will indicate that the chosen peaks cannot refer to one pattern motion, and consequently any velocity calculated from them is meaningless.

If the correlations are very wide in lag relative to their peak location, accurate location of the peaks is difficult. One solution, termed "metacorrelation" is based on the fact that, under the assumption of a Gaussian correlation function, the ACF ( $\rho_A$ ) and CCF ( $\rho_{ij}$ ) have the same shape (e.g. the same width at half peak value). Thus the logarithms of these functions should be the same parabolic function but translated linearly in lag and  $\log(\rho)$ . An accurate value for the shift in lag can be found by lagged correlation between each  $\log(\rho_{ij})$  and  $\log(\rho_A)$ . The position in lag of the peak "metacorrelation" defines the best value of  $t_{\max}$  (the zero-lag value of  $\rho_A$  is omitted in this process because of a possible noise spike here). Another parameter

directly available is the factor by which the original CCF's are reduced from the ACF, which gives the true value of each peak CCF, and bypasses other noise correction methods. However, these latter values were not found to be very accurate in tests.

Meteor trails are a cause of nonstationary data, and since the signal increases in all spaced antennas simultaneously, they cause wide spurious peaks in the CCFs at zero lag. If the data are converted to single-bit amplitudes in blocks, the problem is minimized because the high amplitudes during a block containing a meteor trail will be reduced to have the same significance as those without, in the conversion to binary sequences.

#### VELOCITY CALCULATION

Apparent velocity,  $V_{ap}$ , (which is calculated from just the  $t_{max}$  values has a theoretical disadvantage compared to "true" velocity,  $V_{tr}$ , in which possible pattern elongation and time decay are allowed. In the MF radar case (Saskatoon)  $V_{ap}$  is only -10-20% greater than  $V_{tr}$  below -90 km on the average, and is fairly well behaved except in cases of nonstationary data (e.g. meteor trails) where the values can be extreme; whereas  $V_{tr}$  is virtually always well behaved (although it can be biased by such things as noise level). A calculation of  $V_{tr}$  usually includes that of  $V_{ap}$ , so data can be reworked backwards to find  $V_{ap}$ .

The pattern and time decay parameters in a  $V_{tr}$  analysis are also of interest in terms of the scattering processes. Significant variations have been found in local time as well as seasonally at Adelaide and Saskatoon.