

**A STATISTICAL INFERENCE APPROACH FOR THE
RETRIEVAL OF THE ATMOSPHERIC OZONE PROFILE
FROM SIMULATED SATELLITE MEASUREMENTS OF
SOLAR BACKSCATTERED ULTRAVIOLET RADIATION**

N.L. BONAVIDO¹, C. L. GORDON², R. INGUVA³, G. N. SERAFINO⁴ and R. A. BARNES⁵

ABSTRACT

NASA's Mission to Planet Earth (MTPE) will address important interdisciplinary and environmental issues such as global warming, ozone depletion, deforestation, acid rain and the like with its long term satellite observations of the Earth and with its comprehensive Data and Information System. Extensive sets of satellite observations supporting MTPE will be provided by the Earth Observing System (EOS), while more specific process related observations will be provided by smaller Earth Probes. MTPE will use data from ground and airborne scientific investigations to supplement and validate the global observations obtained from satellite imagery, while the EOS satellites will support interdisciplinary research and model development. This is important for understanding the processes that control the global environment and for improving the prediction of events. In this paper we illustrate the potential for powerful artificial intelligence (AI) techniques when used in the analysis of the formidable problems that exist in the NASA Earth Science programs and of those to be encountered in the future MTPE and EOS programs. These techniques, based on the logical and probabilistic reasoning aspects of plausible inference, strongly emphasize the synergetic relation between data and information. As such they are ideally suited for the analysis of the massive data streams to be provided by both MTPE and EOS.

To demonstrate this, we address both the satellite imagery and model enhancement issues for the problem of ozone profile retrieval through a method based on plausible scientific inferencing. Since in the retrieval problem, the atmospheric ozone profile that is consistent with a given set of measured radiances may not be unique, an optimum statistical method is used to estimate a "best" profile solution from the radiances and from additional apriori information. This method includes a first guess profile and an estimate of its

1. High Performance Computing Branch, Space data and Computing
Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771;
(301) 286-4079. Email: c0nlb@charney.gsfc.nasa.gov

2. SYMBIOTIC TECHNOLOGIES Inc., Riverdale, MD 20737; (301) 779-2449
and Laboratory for Atmospheres, NASA Goddard Space Flight Center,
Greenbelt, MD 20771; (301) 286-0714 Email: PACF::TOMSS\$25A: [Gordon]

3. University of Wyoming, Department of Physics and Astronomy,
Laramie, WY;(307) 766-6150 and SYMBIOTIC TECHNOLOGIES Inc.,
Riverdale, MD 20737; (301) 779-2449. Email: Ramarao@corral.uwyo.edu

4. Goddard Distributed Active Archive Facility, NASA Goddard Space
Flight Center, Greenbelt, MD 20771; (301) 286-3021.
Email: Serafino@eosdata.gsfc.nasa.gov

5. Mantech, Inc., and Laboratory for Hydrospheric Processes, NASA
Goddard Space Flight Center, Greenbelt, MD 20771; (301) 286-0501.
Email: RBarnes@gsfcmail.nasa.gov

variance, the estimated errors in the measurements, and correlations between profile variance and errors of measurement at different levels. The apriori information provides a constraint on which of the solutions consistent with the measured radiances is to be accepted.

A Bayesian analysis of this problem shows that while the data may fully specify the likelihood of a profile, the apriori information is often dismissed as not as fully cogent as the data. In Bayesian estimation, a balance is found between these two in order to ensure that a unique solution can be selected from within the maximum likelihood of feasible solutions. In addition, since the number of levels over which the ozone is distributed is greater than the number of measured radiances, then the problem of inferring the profile from these measured radiances is an ill-posed one. However, the problem is not only ill-posed, it is also nonlinear and since the transfer function is itself dependent on the profile, the information which is passed from the profile plane to the data plane is expressed as a Fredholm integral equation of the first kind. Ozone retrieval thus appears well suited to a statistical inference analysis that encompasses both logical and probability based reasoning.

In this application, a maximum entropy based Bayesian method is introduced which fully utilizes the evidence of prior information and makes logical assignments of numerical values to probabilities from the measured data. A nonlinear transfer function which includes a single scatter model, and a given climatological profile are convolved in order to model twelve solar backscattered ultraviolet (SBUV) radiances. These range from 255.7 to 339.9 nm. The model radiances and the radiative transfer function are then used as input to both the optimum statistical and maximum entropy methods so as to compare the retrieved profiles with the given one. In the maximum entropy approach, both the data values and apriori information are used as constraints on the entropy. This yields a nonlinear equation for the retrieved profile, and the results obtained are seen to compare favorably with the corresponding analysis provided by the standard optimal statistical estimation procedure.

In this environment, we demonstrate the power of inductive inferencing to identify the source of data and then to accurately infer beyond the given data. These considerations are important in the technology of artificial intelligence. In the mammalian brain for example, the inferencing process in which new information or patterns are discovered and from which predictions are made is implicit in the ability called learning. Most raw data reaching the brain is noisy, incomplete and the product of the convolution of several nonlinear sources. How the brain deconvolves these signals and learns from them remains a mystery. We present results of how two powerful methods of inductive inference are used to accomplish this.

INTRODUCTION

Developing a comprehensive understanding of how the Earth functions requires global observations on a sustained, consistent basis for a decade or longer. These observations must provide both a characterization of the state of the whole planet and detailed measurement of its regional variations. They must also enable quantification of the processes that govern the Earth system. Remote sensing of the Earth's environment from space provides the only truly global perspective available. Making the full set of observations goes well beyond the capabilities of any single satellite however, and many of the detailed measurements can only be made in situ. Such a massive network of global observations is planned for the MTPE program. Extensive sets of satellite observations will be provided by both EOS and the Earth Probes in order to support the MTPE.

Among the several mission objectives of the Earth Observing System is included that of understanding the structure, state variables, composition, and dynamics of the atmosphere from the ground to the mesopause. Since remote sensing of the atmospheric composition and profiling by satellites first began, it has become a major technique for the analysis of planetary atmospheres. As a consequence, many sophisticated methods for deriving atmospheric parameters from satellite measured radiation have been developed. To date most methods attempt to deduce a best estimate of the state of the atmosphere from the given measurements, where the intensity and spectral distribution of the latter are assumed to depend on the atmospheric state in a known way. In the problem of the retrieval of the atmospheric ozone profile for example, previous models of the profile and the knowledge of the behavior of radiative transfer are combined with measured data including total ozone, to reach conclusions about the ozone profile. This is a form of deductive analysis in which we classify the solutions as profiles and then require the algorithm to infer the most likely one based on given information. By deductive is meant the special implication of drawing a particular inference from a generalization.

To estimate the profile from the data, we include that phase of plausible reasoning in artificial intelligence (AI) known as inductive inference. By inductive inference we mean the arrival at a conclusion by using available evidence to reason from a part to a whole or from the individual to the universal. A common problem that arises in all data processing is that of how to handle incomplete information. The situation is further complicated if there are several such datasets originating from different sources. What is required to address this is a method that will not only perform multiple source processing of incomplete data, but will also induce inferences from the data. When an inference is made beyond the observational data, a logical relationship between the data and the inference must be expressed. This relation is in a generalized logic, which is not necessarily deductive, and from which inference is neither deductively proved nor disproved from the data. It assesses the support for the inference given the data, but the essential feature is that this support can be of many different degrees. For example, many instances of an event happening, with no exception, in given circumstances, are better evidence than one instance that the event will happen the next time the circumstances occur. This relation between a set of data and a conclusion is called a probability, and the subject is essentially what is called a many valued logic. Generally speaking, probability theory is the system of reasoning applicable in the absence of certainty. This is also known as inductive logic. As such, a probability expresses a degree of reasonable belief. In ordinary logic, a fixed set of postulates is given at the start, and all propositions asserted later are consequences of this set. In probability theory both the data and the proposition considered are subject to alteration, and it is therefore necessary to keep the data explicit. This relation is usually written in the form

$$P(q|p) = a$$

(read the probability of q, given p, is a), where a is the number that expresses the degree of confirmation. A fundamental development of the theory of probability has been provided by Keynes (1929), in which he contends that the above relation expresses an extended logic or a logic of probable inference. It is defined as a relation between a hypothesis and a conclusion, corresponding to the degree of rational belief and limited by the extreme relations of certainty and impossibility. In this sense then, classical deductive logic would reduce to a special case of the more general development since it would fall within the domain of the limiting relations. As a consequence, certainty would be a special case of probability since the latter cannot be based entirely on classical logic. Using this as a basis, Cox (1946) employs the algebra of symbolic logic to derive the rules of probability from

primitive notions which are independent of the frequency concept. In effect he determines the rules or relations of reasonable expectation consistent with symbolic logic.

In terms of both utility and decision modeling, the processing of radiance data for profile estimation represents a vast and realistic class of problems ideally suited for the introduction of inductive inferencing. Generally speaking, data processing can be considered as an operation in which N numbers can be determined from K observations. For $N < K$, any data regarded as exact values, often lead to a mutual inconsistency in the process of determining the best values of N from K . If $N = K$, then a unique solution exists. However, deconvolution under these conditions will lead to unstable solutions if several of the data values contain almost the same information about the conclusions. This in effect leads to an $N > K$ condition and the problem is then ill-posed since there are many conclusions consistent with the same data. This is the case for the profile retrieval. There are two frequently used methods for addressing this. One is model fitting and the other is the addition of data from other sources so as to get $N = K$. In the first case $N < K$ and the problem is changed to one of parameter fitting. Here an answer is obtained whether the assumption is true or false. If the model is not known to be correct, then we have essentially constructed one of the infinity of conclusions that fits the data. When $N = K$, one can assume that the solution is unique and then use one of several standard mathematical approaches to determine it. If however, several of the data points contain the same information, then the problem may again be ill-posed. An example of this can be found in power spectral estimation, which involves a Fourier transformation of data between two canonically conjugate spaces, such as position and momentum. Since there is no data included beyond the range of measurements, these are in effect considered to be zero. The unmeasured Fourier momentum components however are not zero. This assumption which causes a discontinuity in the maximum measured momentum value leads to large oscillations in position space. To overcome this, Fast Fourier transform (FFT) techniques, employ a smoothing of the data by a time domain window. However, the design of these windows are not based on the true spectrum, so that immediate consequences of this are sidelobe leakage in the transfer function of the smoothing window and a limit on the resolution. For a time series of data covering an interval Δt , the energy of the process defining this data will be constrained within this time interval according to the Heisenberg Uncertainty Principle. In addition, the Fourier transform of this time series function confines the energy to a bandwidth $\Delta f \geq (\Delta t)^{-1}$. Consequently, the best resolution attainable is $\Delta f = (\Delta t)^{-1}$. This is because the function is assumed to be zero outside of the interval in which it is given. If the function can be extended or continued in some physically realistic manner, then the spectral frequency resolution will be considerably higher than $(\Delta t)^{-1}$. For a segment of data of a stationary time series which is short compared to the time series itself, the spectral estimation method of Burg (1967) extends this short data sequence to that of a complete series through inductive inferencing employing the maximum entropy principle.

In this paper, we first review the problem of the ozone profile retrieval and then briefly describe Bayesian and maximum entropy concepts. We present results based on a maximum entropy/Bayes algorithm using radiance data generated from a given climatological ozone profile. These results are compared with those of a classical method known as the optimal statistical solution, using this same generated radiance data.

STATEMENT OF THE PROBLEM

In 1934, Gotz et al. (1934) were able to use measurements of diffusely transmitted solar ultraviolet radiation to infer the main features of the atmospheric ozone profile. Since this classic work, there has been extensive analysis on the problems of inferring atmospheric profiles from measurements of solar irradiance backscattered by the atmosphere (Twomey, 1963, 1965, 1966; Twomey and Howell, 1963, 1967; Mateer, 1964, 1965). The possibility of deducing the ozone profile was first suggested by Singer and Wentworth (1957), and the first mathematical examination of the problem was made by Twomey, (1961). He showed that by using a single scatter atmospheric model, and by expressing the mass of ozone above a given pressure level as an explicit function of the atmospheric pressure, the spectral energy distribution of the backscattered radiance was a Laplace transform of the ozone profile. This method was used in some of the earliest work on evaluating measurements of backscattered radiation. The retrieval of the ozone profile from satellite measurements of the solar ultraviolet radiation backscattered by the earth and its atmosphere, is usually divided into two parts: that of the high level profile above 25-30 km, and below this, the inference of the low level profile. In the high level case, a single scatter model is usually adequate to determine backscattered intensity accurately. The corresponding wavelengths here are at 2975 Å and shorter. For wavelengths that penetrate the ozone layer and are backscattered appreciably within the troposphere, multiple scattering calculations are essential and the effects of aerosol scattering as well as cloud and ground reflections become important. A considerable amount of apriori statistical information about the low level ozone profile is available, whereas relatively few reliable data are available for the high level profile.

The inference of atmospheric profiles from radiance measurements usually involves the inversion of an integral equation of the form

$$\int K(x,y)f(x)dx = g(y). \quad (1)$$

The $g(y)$ are the radiance measurements specified at various values of y , $K(x,y)$ is the appropriate kernel, and $f(x)$ is a function of the unknown atmospheric profile. In matrix form, Equation (1) can be written in the form

$$Af = g \quad (2)$$

where A is the matrix that transforms from the $f(x)$ profile plane into the $g(y)$ observation plane and which also allows for the amplitude transmission of differential spatial scales from the $f(x)$ plane to the $g(y)$ plane. Equations such as (1) in which the kernel is also a function of the desired variable, are called Fredholm integral equations of the first kind (Courant and Hilbert, 1953; Fox and Goodwin, 1953; Fox, 1953, 1962, 1964; Phillips, 1964; Tricomi, 1957). In practice, the following approach is used: For a plane parallel atmosphere, the backscattered radiance I , in the satellite nadir direction, with a solar zenith angle θ and a wavelength λ , can be written

$$I(\lambda, \theta) = F_o(\lambda)(3B_\lambda/16\pi)(1+\cos^2\theta) \int_0^1 \exp[-(1+\sec\theta)(\alpha_\lambda X_p + \beta_\lambda p)] dp \quad (3)$$

w h e r e

$F_o(\lambda)$ = extraterrestrial solar irradiance

β_λ = atmospheric scattering coefficient (atm)⁻¹

α_λ = ozone absorption coefficient (atm-cm)⁻¹

and

X_p = amount of ozone above pressure p (atm) in atm-cm.

Equation (3) is considered the starting point for the retrieval of the vertical atmospheric ozone profile. In addition to being ill-posed, it is also ill-conditioned in the sense that there are many solutions which exactly satisfy an integral equation slightly perturbed from the original starting conditions.

In the process of inverting Equation (3), additional apriori constraint information is employed to help reduce the problem to one of estimation. A thorough discussion of this is given by Rodgers (1976). The apriori constraints, are sometimes called 'virtual measurements' since they contain information used in the construction of the profile. These can be derived from the physics, from mathematical restrictions on the solution, or from other independent information. The apriori information used in the optimum statistical method also includes a "first guess" profile obtained from the best available ozone climatology. The latter and its variances and covariances are taken as a function of latitude, time of year, and the total ozone. The radiances that such a profile yields when convolved with a radiative transfer function, is then calculated and the differences between these and the measured or direct radiances are then used to provide a new set of profile values. It is expected that the successively iterated results are more consistent with both the measurements and the first guess profile. The application of this method also requires an assessment of the uncertainty or variance in the measurements and statistical apriori profile information. The former is characterized by errors of measurement and requires covariances in the errors of measurements to determine how dependent the errors at one wavelength are on the errors at another. For the apriori information, the corresponding variances and covariances are obtained in the development of the climatology. A complete description of an inversion algorithm which utilizes the optimum statistical method is given by Fleig et al (1990). This approach proceeds as follows: The backscattered radiation given by Equation (3) is first written in terms of the ratio of backscattered to incident radiation. A single scatter representation for the radiative transfer function is introduced, and the ratio is linearized by expanding in a first order Taylor series about a first guess profile. The problem takes the form of Equation (2) where A is now independent of f. The partial derivatives called the weighting functions, are obtained from the ozone profile of the previous iteration and a solution by inversion is obtained in an iterative fashion using apriori and error information. A "best solution" is arrived at when the rms differences between the measured and estimated radiances is minimized.

THE INTRODUCTION OF PLAUSIBLE INFERENCE

An important concern in data processing is that of identifying the data with its source. This may not always be an easy task. The inclusion of inductive inference methods then become not only an attractive option but also a necessary one (Pearl, 1988). A sound method of plausible inference should ideally consist of a strong interaction between logical and probabilistic reasoning. In the profile retrieval problem for example, one attempts to induce the profile from the data with a minimum of bias. To accomplish this requires two items: a prior probability for each of the possible classifications, and the values of the

conditional probabilities of the attributes that define the classes. This is the probability of seeing the data given the class and is called the likelihood of the data. Cox (1979) and Horvitz (1986) provide a thorough and in depth discussion on this. They maintain that with the satisfaction of a certain set of specific conditions, the standard "axioms" of probability theory including Bayes' Theorem (Bayes, 1763; Jeffries, 1983) will then follow logically and can be uniquely defined.

Fundamentally, Bayes' Theorem means calculating the posterior conditional probability

$$P(H_i|DI) = P(H_i|I) P(D|H_i,I) / P(D|I) \quad (4)$$

where H_i represents an hypothesis in a sequence of hypotheses ($H_1, \dots, H_i, \dots, H_n$), which form a complete set and whose truth one wishes to judge. D is a set of data, and I is whatever prior information one has in addition to the data. The inference can then be summarized such that if H_i is the desired profile, then the best estimate of H_i , in light of the data and any apriori information, is given by that profile which maximizes this posterior probability. Bayes' Theorem thus relates this probability that we require to the two others, one of which can be computed directly. Here $P(H_i|I)$ is the prior probability and represents the state of knowledge (or ignorance) about the profile before there is any data. This prior state of knowledge is modified by the data through the likelihood function or conditional probability $P(D|H_i,I)$. This quantity indicates how likely it is that a particular data set would have been obtained from a given (trial) hypothesis. In a typical classification problem, the prior and likelihood terms will compete as to the number of classes present with the likelihood preferring the largest number possible and the prior preferring the least. The conditions which provide a number acceptable to both, will also yield the highest value of the posterior probability. If an experiment is performed and new data D occurs, then a reevaluation is required of the hypothesis H_i in order to calculate the new conditional probability (the left hand side of Equation (4)). With the continued occurrence of data D in repeated experiments, we tend to believe more in the hypothesis H_i at the expense of believing less in the others. The prior probability can be "well-behaved" as is the case whenever this function possesses a single maximum. It can also be "badly behaved" in the sense of having many local maxima. This is usually the case when the data and the desired variable are nonlinearly related. In such circumstances, techniques involving simulated annealing methods are sometimes used to avoid producing local subsidiary solutions (Kirkpatrick, Gelat & Vecchi, 1983). These are usually of little help whenever many almost equally probable solutions are present.

Since the functional form for the likelihood of the data depends essentially on the nature of the source producing the data, then the posterior probability will inherit much of this complex topology. An example of this is found in the restoration of the blurred and noisy image of the binary stellar system R Aquarii, provided by the Hubble Space Telescope Faint Object Camera, (Bonavito et al., 1993). Here the image suffered from both spherical aberration and detector saturation and was characterized by sharp peaks of intensity within data cells immersed in a dim background. Datasets such as these are subject to noise governed by Poisson statistics which are then modeled in the likelihood function.

For complex systems requiring extensive calculations, Bayesian networks show some promising developments (Pearl, 1988; Charniak & McDermott, 1985). To determine the posterior distribution of a Bayesian network, one must specify the prior probabilities of what are termed root nodes (or nodes with no predecessors) on an AI graph. It is also necessary to specify the conditional probabilities or likelihood of all of the nonroot nodes

(evidence or data), given all possible combinations of their direct predecessors. Bayesian networks allow one to calculate the conditional probabilities of the nodes in the network given that the values of some of the nodes have been observed. They are calculated from a small set of probabilities relating only to neighboring nodes. The nodes can be considered as random variables representing various states of affairs. In realistic cases, the networks may consist of thousands of nodes which are evaluated many times as new evidence comes in. What changes then is the conditional probability of the nodes given the new data. The ability of the networks to greatly reduce the complete specification of a probability distribution in complex systems using built-in independence assumptions, now makes extensive Bayesian analyses realizable.

THE MAXIMUM ENTROPY METHOD

Maximum entropy has its roots in the work of Boltzmann (1877) and Gibbs (1875) near the latter part of the last century and in the work of Shannon (1948). It has to do with drawing inferences from incomplete information. Fundamentally, it states that any inferences made concerning the outcome of any natural process should be based upon the probability distribution which has the maximum entropy permitted by the data taken during observation of the process. Here the data is defined as ensemble averages,

$$d_k = \sum_{j=1}^n P_j A_{kj}, \quad 1 \leq k \leq m, \quad (5)$$

where A_{kj} defines the nature or physics underlying the measured quantities, and the P_j , the distribution upon which the ensemble averages are imposed as constraints. Then as shown by Gibbs (1875) and Jaynes (1957), using the the method of Lagrange multipliers, with the partition function,

$$Z(\lambda_1, \dots, \lambda_m) = \sum_{j=1}^n \exp\left(- \sum_{k=1}^m \lambda_k A_{kj}\right) \quad (6)$$

the maximum entropy distribution is,

$$P_j = \frac{1}{Z(\lambda_1, \dots, \lambda_m)} \exp\left(- \sum_{k=1}^m \lambda_k A_{kj}\right), \quad 1 \leq j \leq n. \quad (7)$$

The Lagrange multipliers λ_k are obtained from

$$\frac{\partial \ln Z}{\partial \lambda_k} + d_k = 0, \quad 1 \leq k \leq m, \quad (8)$$

a set of m simultaneous equations for m unknowns. Any other distributions allowed by the constraints (5), will necessarily have entropy values less than those determined by Equation (7). The fact that P_j is a positive quantity has important implications in all areas of signal processing. There are as many Lagrange multipliers as there are equations of constraint, and these constitute the disposable parameters of the minimally prejudiced probability distribution. They are to be so adjusted as to satisfy the given data. From this, one may conclude that maximum entropy is the appropriate method to 'reason' from the microscopic to the macroscopic. Thus, if one wishes to consider the expectation values for the measured data values given by Equation (5), as entities for which the sum of the probabilities is equal to one, then the corresponding measured values can be said to introduce an element of logical reasoning into the problem of plausible inference. In this sense, they help to determine the consequences of the model for the given constraints. On the other hand, one can also consider that probabilistic reasoning, which enters through Equation (7), is required to interpret the plausibility of the model.

ESTIMATION OF THE ATMOSPHERIC OZONE PROFILE

The problem that we address in this paper is defined as follows: We convolve a known ozone profile at a specified latitude, time of day and solar zenith angle, with a given radiative transfer function using twelve ozone absorption and twelve atmospheric scattering coefficients. This produces twelve model SBUV radiance data values. Using these simulated data values together with the SBUV estimated total ozone and the radiative transfer function, the task is to retrieve the above known (Given) ozone profile used in the convolution.

Let us summarize some of the key issues pertaining to the retrieval problem. We first note that there are more levels over which to distribute the total ozone than there are measured data values (the ill-posed problem). The transfer function is also non-linear and is itself a function of the desired profile. This gives rise to an expression for the backscattered radiation that is in the form of a Fredholm integral equation of the first kind. These integral equations are very difficult to solve and often times unwarranted assumptions are imposed in order to handle them. The problem is also ill-conditioned in that there are many possible solutions which exactly satisfy this integral equation whenever the original starting conditions are slightly perturbed.

The problem is then formulated in terms of the atmospheric pressure. This is possible since the altitude above the surface (except for minor local barometric fluctuations) and the ozone amount distribution, are each be expressed parametrically as a function of atmospheric pressure. It is also useful to choose atmospheric pressure as the independent variable, since atmospheric pressure, not altitude, has a direct influence upon the scattering of the ultraviolet solar radiation.

Measurements of backscattered ultraviolet solar radiation are made at a small number, m of wavelengths, so that in order to facilitate calculation, the atmosphere is divided into n layers, where n is greater than m . A large number of layers is sometimes used to obtain a smooth curve representing the amount of ozone in each layer. In what follows, x_j represents the amount in the j th layer and T is the total amount of ozone in all of the layers.

To adapt the maximum entropy method to the problem of profile retrieval requires the identification of the probability distribution (Equation (7)), with an appropriate profile parameter such as the fraction of total ozone received at a particular level,

$$f_j = x_j / T. \quad (9)$$

From this, the distribution can be written as

$$P_j = \frac{f_j}{\sum_{j=1}^n f_j} \quad (10)$$

As a consequence, P_j can then be replaced by f_j where $f_j \geq 0$ at each level. This positivity constraint is guaranteed by the exponential in the maximum entropy solution.

It is convenient to describe the observed radiances of Equation (3) in terms of a quantity Q_λ defined as the ratio of incident to backscattered radiation

$$Q_\lambda = \int_0^{p_N} \exp(-[(\hat{\alpha}_\lambda X_p + \hat{\beta}_\lambda p)] dp) \quad (11)$$

where $\hat{\alpha}_\lambda = \alpha_\lambda (1 + \sec \theta)$ and $\hat{\beta}_\lambda = \beta_\lambda (1 + \sec \theta)$. Equation (11) is of the form of a Fredholm integral equation of the first kind and an approximate maximum entropy solution for this type of equation as well as those of the second kind and the Wiener-Hopf type have been developed by Mead (1986) and Papanicolaou (1984). In their approach, generalized moments are introduced into the integral equation and the problem is converted into an equivalent one in which the informational entropy is maximized using these moments as constraints. Rather than utilizing this approach however, we proceed as follows: The integral in Equation (11) is discretized by dividing the atmosphere into n layers

$$Q_\lambda = \sum_{j=1}^n A_{\lambda j} g_{j\lambda} \quad (12a)$$

$$\text{where } A_{\lambda j} = [\exp(-\hat{\beta}_\lambda p_j)] W_j \quad (12b)$$

$$\begin{aligned} W_j &= 0.5 (\Delta p_j + \Delta p_{j+1}), \text{ for } j = 1, 2, \dots, n-1 \\ &= 0.5 \Delta p_n, \text{ for } j = n \end{aligned} \quad (12c)$$

$$\text{and } g_{j\lambda} = \exp(-\hat{\alpha}_\lambda \sum_{k=1}^j x_k). \quad (12d)$$

Here x_k is the ozone amount in the k th layer. From this, one can then define

$$\Delta p_j = p_j - p_{j-1}, \quad (12e)$$

where p_j is the pressure at the bottom of the j th layer. Equation (12a) is a nonlinear equation for the ozone layer distribution $\{x_1, x_2, x_3, \dots, x_n\}$. Maximizing the informational entropy subject to the normalization

$$\sum_{j=1}^n f_j = 1 \quad (13)$$

and the m constraints Q_λ , where $\lambda = 1, 2, \dots, m$, yields

$$f_j = 1/Z \exp(\Gamma_j). \quad (14)$$

$$\text{Here } Z = \sum_j \exp(\Gamma_j), \quad \Gamma_j = \sum_{k=1}^N \sum_{\lambda=1}^M C_\lambda \gamma_\lambda A_{\lambda k} g_{k\lambda} \quad \text{and} \quad C_\lambda = \hat{\alpha}_\lambda T.$$

The γ_λ are the Lagrange multipliers which couple the constraints Q_λ .

RESULTS AND DISCUSSION

All of the data used in this problem including that which comprise the curves shown as Given and Guess in Figure 1, and that which was used to evaluate the maximum entropy and optimum statistical technique profiles, were provided by the Atmospheric Chemistry Branch of the Laboratory for Atmospheres at the Goddard Space Flight Center. The Rayleigh scattering and ozone absorption coefficients which define the spectroscopic character of this particular radiative transfer function are shown in Table 1. The left hand column are the twelve wavelength values for which twelve corresponding data values of Q_λ given by Equation (12a) were produced during the convolution process. The value for the solar zenith angle was taken to be 69 degrees.

Table 1
Absorption and Scattering Coefficients

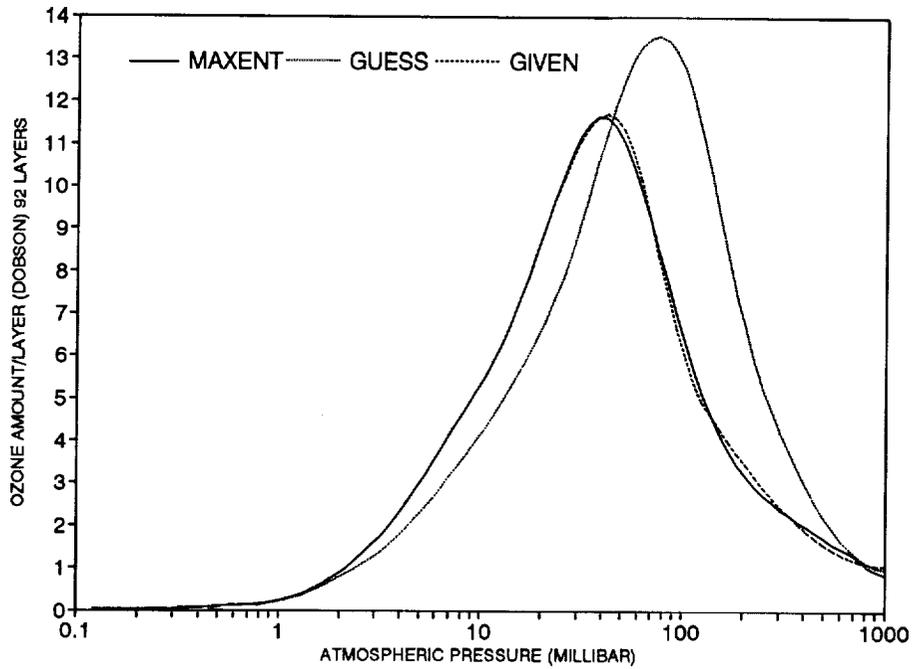
Wavelength (nm)	Ozone Absorption Coefficient (atm-cm) ⁻¹	Rayleigh Scattering Coefficient (atm) ⁻¹
255.7	309.7	2.4573
273.6	169.9	1.8131
283.1	79.88	1.5660
287.7	48.33	1.4597
292.3	27.82	1.3627
297.6	13.66	1.2605
302.0	7.462	1.1831
305.9	4.281	1.1194
312.9	1.632	1.0198
317.6	0.8684	0.9527
331.3	0.1397	0.7956
339.9	0.0248	0.6864

The ozone distribution for the curve labeled Given was obtained from twelve larger layers of ozone known as Umkehr layers. These Umkehr layers are in turn derived from an algorithm which is based on climatology information. The twelve values are then cubic spline interpolated to 92 layers to yield what is shown as the Given curves of Figures 1(a) and (b). These 92 values were then used to obtain the convolution of a single scattering radiative transfer function given by Equation (3). The total ozone was obtained by the summation of the ozone amount at each of the 92 levels. To convert to Dobson units, these are multiplied by 1000. The curves shown as the Guess on Figures 1(a) and (b), are obtained by changing the day number, latitude and total ozone in the above.

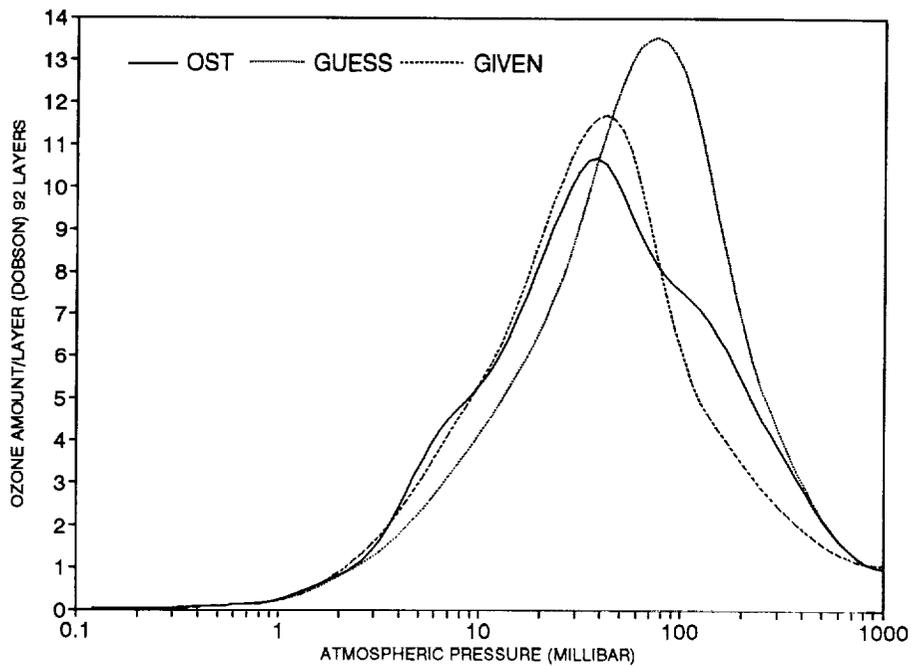
Figure 1(a) shows the maximum entropy retrieved profile. It is clear from these results that this inversion is very close to the Given profile, that is, the one used as the known profile in this example. This agreement is almost exact at all pressure values from below 1 mb up to 1 atmosphere. Figure 1(b) depicts the profile retrieved by the optimum statistical technique for this same Given profile.

This example has allowed us to demonstrate the power of inductive inference methods not only to identify correctly the most likely source of a dataset but also to accurately "predict" new information. In Bayesian analysis the sample probabilities are used to induce an hypothesis which most likely will identify with the data source. In this way, Bayesian statistics involves learning something about the assumptions by looking at the results. This provides a quantitative way to evaluate the probabilities of different assumptions, given the data. This is important in science for example, where there are often competing hypotheses for the explanation of some natural phenomenon. Going back into the unknown, using the observations, is what characterizes Bayesian statistics. In this sense it uses data to test hypotheses. Maximum entropy on the other hand, uses the model identified with the data source to make inferences about the data samples. This cannot be done in classical statistics.

The process of updating knowledge by introducing new data is a basic one in the animal ability called learning. This is complicated by the fact that raw data reaching the mammalian brain is more likely than not the result of the convolution of several nonlinear sources which may have generated datasets that are noisy and incomplete as well.



(a)



(b)

Figure 1: Ozone retrieval by (a) Maximum Entropy and (b) Optimum Statistical Technique (Courtesy of Symbiotic Technologies, Inc.)

The ability to deconvolve these signals and learn from them has fascinated experimental psychologists for more than a century. A recent trend by members of this community has been to pursue the idea that individuals solve particular kinds of problems by making specific inferences (deductive) using rough guidelines that keep track of conclusions compatible with the information at hand and along with relevant prior knowledge. The best fit between the premises of a problem and the acceptable conclusion is judged to be plausible. Psychologists have also focused on ideas which may be involved in forming decisions out of incomplete or ambiguous pieces of information. It is under these conditions however that the human brain often falls prey to what is called "cognitive illusions." More recently, the ability "to understand" has occupied the attention of scientists and engineers engaged in the field of machine learning. The studies surrounding learning in the brain has split along the two lines of thought called behaviorism and cognitivism. These center about the question of whether learning is a matter of behavioral patterning by reinforcement or the storage and use of knowledge. The early behaviorists considered learning as automatic and machinelike. They observed that if a particular response to a particular stimulus pays off for an organism, then the response is likely to be repeated and the probability that the response will be further repeated will be increased by further rewards. It holds that behavior begins as essentially random activity, but connections are strengthened between stimuli and response when the latter are followed by a satisfying result. Known as reinforcement, this is said to strengthen a response, thereby making it more probable. Complementing this concept is that of cognitivism which holds that from the process of reinforcement, information is retained and is confirmed or not by experience, resulting in learning. With the addition of information to random neural systems and with the development of expectations about how certain goals can be achieved, both perspectives can be viewed as the psychological analogues to self-organization. In a very similar fashion, this can also be viewed as a Bayesian description of learning.

REFERENCES

- Bayes, T. (1763). *Philosophical Transactions of the Royal Society of London*, 330-418.
- Boltzmann, L. (1877). *Über die Beziehung Zwischen dem Zweiten Hauptsatzender Mechanischen Warmtheorie und der Wahrscheinlichkeitsrechnung, Respective der Satzenuber das Warmgleichgewicht*. *Wien. Bericht*, 76, 373-395.
- Bonavito, N. L., Dorband, J. E., & Busse, T. (October, 1993). *Maximum Entropy Restoration of Blurred and Oversaturated Hubble Space Telescope Imagery*. *Journal of Applied Optics*, 32, No. 29, 5768-5774.
- Burg, J.P., (1967). *Maximum Entropy Spectral Analysis*. Proc. 37th Meeting, Soc. Explor. Geophys. Oklahoma City, OK.
- Charniak, E., & McDermott, D. (1985). *Introduction to Artificial Intelligence*. Reading, MA.: Addison-Wesley.
- Courant, R., & Hilbert, D. (1953). *Methods of Mathematical Physics*. Vol. I. New York and London: Interscience.
- Cox, R. T. (1946). *Probability, Frequency and Reasonable Expectation*, *American Journal of Physics*, 9, 1-13.

- Cox, R. T. (1979). *Of Inference and Inquiry-an Essay in Inductive Logic: The Maximum Entropy Formalism*. L. Levine and M. Tribus (Editors). Cambridge, MA: MIT Press.
- Fleig, A. J., McPeters, R. D., Bhartia, P. K., Schlesinger, B. M., Cebula, P., Klenk, K. F., Taylor, S. L., & D. F. Heath. (January 1990). *Nimbus 7 Solar Backscatter Ultraviolet (SBUV) Ozone Products User's Guide*, NASA Reference Publication 1234.
- Fox, L. (1964). *Introduction to Numerical Linear Algebra*. Oxford: Clarendon Press.
- Fox, L. (1962). *Numerical Solution of Ordinary and Partial Differential Equations*. Oxford: Pergamon.
- Fox, L., & Goodwin, E. T. (1953). The Numerical Solution of Non-singular Linear Integral Equations. *Philosophical Transactions of the Royal Society*, A 241, 501.
- Gibbs, J. W. (1875). *Heterogeneous Equilibrium*. New Haven: Conn. Academy of Science. Reprinted (New York: Longman, Green and Co., 1928); reprinted (New York: Dover, 1961).
- Gotz, F. W. P., Meetham, A. R., & Dobson, G. M. B. (1934). The Vertical Distribution of Ozone in the Atmosphere. *Proceedings of the Royal Society of London*, Ser. A. 145, 416-446.
- Horvitz, E. J., Heckermann, D. E., & Langlotz, C. P. (1986). Framework for Comparing Alternate Formalisms for Plausible Reasoning. *Proceedings of the 5th National Conference on Artificial Intelligence*. Philadelphia, PA. 210-214.
- Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106, 620-630 and 108, 171-190.
- Jaynes, E. T. (1982). On The Rational of Maximum Entropy Methods. *Proceedings , IEEE*, 70, No. 9, 939-952.
- Jeffreys, H. (1983). *Theory of Probability*. Oxford University Press, Fourth Edition.
- Keynes, J. M. (1929). *A Treatise on Probability*, London: Macmillan.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). *Science*, 220, 671-680.
- Mateer, C. L. (1964). A Study of the Information Content of Umkehr Observations. Phd. Thesis, University of Michigan.
- Mateer, C. L. (1965). On the Information Content- of Umkehr Observations. *Journal of the Atmospheric Sciences*, 22, 370-381.
- Mead, L. R. (1986). Approximate Solution of Fredholm Integral Equations by the Maximum Entropy Method. *Journal of Mathematical Physics*, 27, 2903-2907.
- Mead, L. R., & Papanicolaou, N. (1984). *Journal of Mathematical Physics*, 25, 2404.

- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 2nd Ed. San Mateo, CA.
- Phillips, D. L. (1962). A Technique for the Numerical Solution of Certain Integral Equations of the First Kind. *Journal of the Association of Computing Machinery*, 9, 84.
- Rodgers, C. D. (1976). Retrieval of Atmospheric Temperature and Composition from Remote Measurements of Thermal Radiation. *Reviews of Geophysics and Space Physics*, 14, 609-624.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 623-656.
- Singer, S. F., & Wentworth, R. C. (1957). A Method for the Determination of Vertical Ozone Distribution from a Satellite. *Journal of Geophysical Research*, 62, 299-308.
- Tricomi, F. G. (1957). *Integral Equations*. New York and London: Interscience.
- Twomey, S. (1961). On the Deduction of the Vertical Distribution of Ozone by Ultraviolet Spectral Measurements from a Satellite. *Journal of Geophysical Review*, 66, 2153-2162.
- Twomey, S. (1963). On The Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature. *Journal of the Association of Computing Machinery*, 10, 97-101.
- Twomey, S. (1965). The Application of Numerical Filtering to the Solution of Integral Equations Encountered in Indirect Sensing Measurements. *Journal of the Franklin Institute*, 279, 95-109.
- Twomey, S. (1966). Indirect Measurements of Atmospheric Temperature Profiles from Satellites. II. Mathematical Aspects of the Inversion Problem. *Monthly Weather Review*, 94, 363-366.
- Twomey, S., & Howell, H. B. (1963). A Discussion of Indirect Sounding Methods with Special Reference to the Deduction of Vertical Ozone Distribution from Light Scattering Measurement. *Monthly Weather Review*, 91, 659-664.
- Twomey, S., & Howell, H. B. (1967). Some Aspects of the Optical Estimation of Microstructure in Fog and Cloud. *Journal of Applied Optics*, 6, 2125-2131.
- U. S. *Standard Atmosphere*. (1976). National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, United States Air Force. 39.