REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	5Dec97	Final Report 7May97 - 7Nov97

4. TITLE AND SUBTITLE
Retrieval Algorithms for Atmosphere Data Assimilation

5. FUNDING NUMBERS
C N00024-97-C-4127

6. AUTHOR(S)
Dr. Jerry D. Lumpe

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Computational Physics, Inc.
2750 Prosperity Avenue, Suite 600
Fairfax, VA 22031

8. PERFORMING ORGANIZATION REPORT NUMBER
Job 5086

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
TPOC: Dr. Richard Bevilacqua, Code 7220
4555 Overlook Avenue SW
Washington, DC 20375-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Report developed under SBIR contract. This report details the development of algorithms for the purpose of assimilating multiple satellite remote sensing data sets of important geophysical parameters into instrument-independent three-dimensional gridded distributions. The assimilation problem has been formulated and solved as a general nonlinear retrieval problem, using the theory of optimal estimation. A detailed description of the method, and the specific structures resulting from its application to data assimilation, are provided. The algorithms have been tested on simulated satellite data sets for the specific problem of creating global ozone mixing ratio distributions from assimilation of satellite limb-viewing occultation and emission data. The results of these simulations clearly demonstrate the technical feasibility of the proposed approach. The potential applications of a general, rigorous data assimilation algorithm are widespread because of the increasing dependence on, and sophistication of, satellite remote sensing data in both the defense and civilian sectors. Examples include the suite of polar orbiting satellites operated by DMSP and NOAA which provide climatological data for operational weather prediction, multi-platform scientific missions such as NASA's planned EOS program, and commercial earth remote sensing programs such as LANDSAT and the French SPOT program.

14. SUBJECT TERMS
SBIR Report Remote Sensing Retrieval Algorithms OOAM / POAM III
Satellite Data Assimilation Optimal Estimation Global Data Sets

15. NUMBER OF PAGES
22

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Retrieval Algorithm for Atmospheric Data Assimilation

Final Report

Contract No. N00024-97-C-4127, SBIR Topic N97-077

Prepared for:

Naval Sea Systems Command
2531 Jefferson Davis Highway
Arlington, VA 22242-5160

Prepared by:

Jerry D. Lumpe

Computational Physics, Inc.
2750 Prosperity Ave. Ste. 600
Fairfax, VA 22031

December 5, 1997

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.
- Pages with background color or light colored printing.
- Pages with small type or poor printing; and or
- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☒ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

Table of Contents

1. Introduction and Summary of Phase I Objectives.

The overall objective of the Phase I effort was to develop and test a working prototype algorithm which can be applied to the problem of satellite data assimilation. For our purposes the data assimilation problem may be defined as follows. Given multiple, independent satellite data sets of a geophysical parameter, for example temperature or the concentration of a particular trace gas, assimilate these data into a unified instrument independent data set on a predefined spatial grid. The resulting three dimensional spatial distribution should be constructed from a statistically optimum combination of the input measurements, accounting for their respective uncertainties and instrument characteristics, as well as known atmospheric state parameters and a priori estimates of the distribution. The specific objectives outlined in the Phase I proposal were as follows: 1) formulate the data assimilation problem in terms of a general retrieval problem; 2) develop the necessary computer programs to implement this formulation numerically, and 3) test the resulting algorithms in realistic simulations incorporating specific instruments and platforms of interest to the sponsor.

As described in this report, we have successfully formulated the data assimilation problem into an appropriate retrieval problem and developed the necessary computer codes to implement these algorithms. The prototype algorithms have been tested successfully on simulated satellite data sets. A detailed description of the technical approach used, along with a discussion of the simulation results, are presented in the following sections. These initial results clearly demonstrate the technical feasibility of the approach outlined in the Phase I proposal.

2. Technical Approach

The approach to the data assimilation problem which has been developed and applied in Phase I utilizes the method of optimal estimation, which is a constrained linear inversion technique. Before describing the method in detail, it is important to clearly state the problem to be solved and to define some basic quantities. The objective of the assimilation is to determine the average spatial distribution of a geophysical variable, x, over a fixed time interval, for example one week or one month. All available data during this time interval are binned together and assumed to be equally valid for that time frame. In other words we are concerned here with a static three-dimensional assimilation problem in which there is no explicit (or implicit) time dependence. Given a finite number of measurements, which are distributed on some irregular observation grid determined by both the satellite orbit parameters and the measurement technique being utilized, we retrieve an optimum distribution on a fixed three-dimensional retrieval grid.

Let y represent a column vector containing observations at all the measurement grid points $\mathbf{r}_{dat}^{i}$ $(i = 1, NY)$. The corresponding covariance matrix of the measurements is

denoted by S_y. Similarly, let x_a represent a column vector of a priori estimates of the true distribution, defined on the retrieval grid points $\mathbf{r}'_{ret}$ $(i = 1, NX)$. The a priori covariance matrix is denoted by S_a. If both the measurement and a priori errors are assumed to be normally distributed then a maximum likelihood estimate of the true distribution, $\hat{x}$, is found by minimization of the scalar cost function

$$\Phi = (\hat{x} - x_a)^T S_a^{-1} (\hat{x} - x_a) + (y - F(\hat{x}))^T S_y^{-1} (y - F(\hat{x})) \tag{1}$$

Here F is the forward operator, which maps the atmospheric state vector x into the measurement space vector y:

$$y = F(x) \tag{2}$$

This operator may be either linear or nonlinear. If it is linear then $y = K \cdot x$ and it is straightforward to show that the functional (1) is minimized if

$$\hat{x} = x_a + S_a K^T \left(K S_a K^T + S_y\right)^{-1} (y - K x_a) \tag{3}$$

If the forward operator is nonlinear then to obtain $\hat{x}$ we iterate towards the solution, linearizing about the current estimate, x^n, each time:

$$F(\hat{x}) \approx F(x_n) + K_n \cdot (\hat{x} - x_n) \tag{4}$$

where the Jacobian matrix is defined as

$$K_n \equiv \left.\frac{\partial F}{\partial x}\right|_{x=x_n} \tag{5}$$

The final solution is then obtained by iterating the following equation to convergence:

$$x_{n+1} = x_a + S_a K_n^T \left(K_n S_a K_n^T + S_y\right)^{-1} \left[y - y_n - K_n (x_a - x_n)\right] \tag{6}$$

Equation (6) is appropriate when the data space is smaller than the retrieval space, i.e., $NY < NX$. An alternative form, which is algebraically equivalent but more computationally efficient for the case $NY > NX$, is given by

$$x_{n+1} = x_a + \left(K_n^T S_y^{-1} K_n + S_a^{-1}\right)^{-1} K_n^T S_y^{-1} \left[y - y_n - K_n\left(x_a - x_n\right)\right] \tag{7}$$

A number of useful convergence criteria may be defined to determine when the iteration is terminated. In the work described here we have generally adopted the criteria that $K_n \cdot \left[x_{n+1} - x_n\right] \leq \sigma_y$ for all points, where σ_y^i is the standard deviation of the i^{th} data point. This criterion is simply equivalent to requiring that the retrieval fit the data to within its expected uncertainty.

The optimal estimation formalism described in this section has been implemented in a very generic and modular form in the FORTRAN code "OPT", which has been developed over the past several years by Dr. Lumpe and others at CPI. These codes are very general and perform either the linear or nonlinear solution, depending on the nature of the forward function appropriate to the problem of interest. They also have some optimization capability, in that the algorithms determine automatically which form of the iteration equation (Equations (6) and (7)) is most efficient to use for the nonlinear case. To apply the algorithms to a particular problem the user needs only to define the appropriate forward function, as well as provide the necessary interface codes to define the data and retrieval space vectors, y and x, and their respective covariance matrices. In the following section we describe how the OPT algorithm has been adapted to the data assimilation problem and show specific examples using simulated data for two different scenarios of satellite limb sounding measurements of ozone.

3. Application to Data Assimilation and Initial Results.

In the Phase I effort the necessary FORTRAN codes have been developed to apply the optimal estimation technique, in the form of the OPT algorithm, to the data assimilation problem. These codes have been developed in as general a manner as possible. The assimilation may be performed on any geophysical quantity of interest, since the physical dimensions of the variables are unimportant and the codes themselves work with variables which are normalized to the a priori vector and therefore are always of order unity. The algorithm automatically allows for the inclusion of an arbitrary number of independent satellite data sets, each with different horizontal and vertical sampling grids, vertical resolution, and measurement error characteristics. The number of data sets, as well as the corresponding file names containing the data from each instrument, are specified in a master input file. The primary assumption that has been made in the prototype algorithms is that the x and y variables are physically identical. In other words the satellite data are assumed to be retrieval profiles rather than measured

radiances. The forward operator F in this case is then essentially an interpolation and smoothing operator, as discussed in more detail below.

3.1 Structure of Data and Retrieval Spaces

The first step in the assimilation is to construct the desired retrieval grid $\mathbf{r}_{ret} \equiv (z,\theta,\varphi)_{ret}$ where z is altitude, θ is latitude, and φ is longitude. Generally these will be uniform grids in all three dimensions, although that is not explicitly assumed in the codes and therefore is not a requirement. The retrieval vector x simply contains the values of the retrieved distribution at each of the grid points. In principle the ordering of this vector is arbitrary, however the following standard ordering has been implemented in the algorithms because it provides the most efficient FORTRAN indexing, and therefore the fastest run time, in practice:

$$\hat{x} = \left[x(z_1,\theta_1,\varphi_1) \ldots x(z_{nz},\theta_1,\varphi_1) \ldots x(z_1,\theta_{nlat},\varphi_{nlon}) \ldots x(z_{nz},\theta_{nlat},\varphi_{nlon})\right] \tag{8}$$

In other words, it is constructed by successively stringing together the complete altitude profile at each of the consecutive latitude/longitude grid points. Note that the total dimension of the retrieval space is given by $NX = nz \cdot nlat \cdot nlon$, where nz, $nlat$, and $nlon$ are the number of altitude, latitude and longitude grid points, respectively. The explicit grids used in the simulations are discussed in more detail below.

The structure of the a priori covariance matrix, S_a, is determined by three parameters, all of which are read in from the primary input file and are therefore easily changed. The diagonal elements, which determine the variance assigned to the a priori profile at each grid point, are simply taken to be a constant fraction of the a priori at all points:

$$\sqrt{S_a^{ii}} = C\, x_a^i \tag{9}$$

For the off-diagonal elements we have introduced spatial correlations in both the vertical and horizontal directions. These correlation terms are assumed to have the following Gaussian form:

$$S_a^{i,j} = \sqrt{S_a^{i,i} S_a^{j,j}}\, \exp\left(-\left(z^i - z^j\right)^2 / H_{vert}^2\right) \exp\left(-d^2 / H_{hor}^2\right) \tag{10}$$

where d is the horizontal distance between the latitude/longitude grid points $(\theta,\varphi)_{ret}^i$ and $(\theta,\varphi)_{ret}^j$. Obviously if the two grid points are either on the same altitude surface, or

at different altitudes but on the same latitude/longitude point, then only one or the other of the exponential terms in (10) differs from unity. The scale factor C and the vertical and horizontal correlation lengths, H_{vert} and H_{hor}, are important parameters which can be tuned to optimize the amount of smoothing constraint used in the retrievals. They also determine the relative amount of information content in the retrieval coming from the data versus the a priori. Quantifying this information content will be a major focus of the Phase II effort.

The data vector is constructed in a similar way to the retrieval vector, by simply stringing together consecutive profile measurements. As mentioned above the codes are configured to handle an arbitrary number, M, of independent data sets from different instruments. This parameter, and the corresponding data file names, is read in from the main input file. One input data file contains the measured profiles and geographical locations of all measurements for a given instrument within the desired time frame. The entire data file is read in and then filtered to retain only those measurements that fall within the bounds of the retrieval grid space. Obviously if the goal is to perform a complete global assimilation then the entire data set will be retained. However in many instances, as in the simulations discussed below, we are only interested in determining a distribution over a limited geographical region. Because the elements of the Jacobian matrix K_n will be zero for data grid points outside the bounds of the retrieval grid these measurements are filtered out at the beginning.

We assume that each instrument i contributes a total of Γ_i separate data profiles and we use the index pair (i,j) to denote the j^{th} measurement of the i^{th} instrument. With each data profile is associated a measurement location, specified by a latitude/longitude pair $(\theta,\varphi)_{dat}^{i,j}$, and an altitude grid $\{z\}_{dat}^{i,j}$. This altitude grid is allowed to be different for each measurement event, even for a given instrument, as this will in general be the case with real satellite data sets. We define the number of vertical levels for the event (i,j) to be $L^{i,j}$. The explicit structure of the data vector is then taken to be the following:

$$y = \left[\left\{y^{1,1}[(z,\theta,\varphi)_{dat}^{1,1}] \ldots y^{1,\Gamma_1}[(z,\theta,\varphi)_{dat}^{1,\Gamma_1}] \ldots y^{M,1}[(z,\theta,\varphi)_{dat}^{M,1}] \ldots y^{M,\Gamma_{M1}}[(z,\theta,\varphi)_{dat}^{M,\Gamma_{M1}}]\right]\right. \tag{11}$$

The total dimension of this vector, which defines the data space, is given by $NY = \sum_{i=1}^{M}\sum_{j=1}^{\Gamma_i} L^{i,j}$.

The construction of the data covariance matrix, S_y, is similar to the a priori covariance defined in Equations (9-10), but with several important differences. First the diagonal elements, corresponding to the variance of the measurements, are given by the actual measurement error profiles for each instrument rather than from the simple

constant relative error assumed in (9). Error profiles are read in from the simulated data files just as individual experimenters would provide them when using real satellite data sets. Second, while we include off-diagonal elements corresponding to vertical error correlations for individual measurement profiles, we assume that there are no correlated errors between distinct measurement events for a given data set, and certainly not between distinct data sets. Therefore S_y^{ij} will contain a term similar to the first exponential factor in Equation (10) but no term equivalent to the second factor. The vertical correlation length for a given measurement profile is taken to be the effective vertical resolution of the instrument, as defined by the full-width-half-maximum (FWHM) of the instrument averaging kernel.

3.2 Forward Model

As we mentioned previously, the prototype assimilation algorithms developed under Phase I explicitly assume that the x and y spaces correspond to physically identical variables. The forward function therefore simply maps the background geophysical distribution, defined on the uniform retrieval grid $(z,\theta,\varphi)_{ret}$, into the individual measurement profiles $y^{i,j}\left[(z,\theta,\varphi)_{dat}^{i,j}\right]$. This mapping involves two separate steps. The first step is the horizontal interpolation from the retrieval latitude/longitude grid to the individual measurement locations. This step is performed independently on each level in the retrieval altitude grid. It may be represented symbolically as follows:

$$x\left[z_{ret},(\theta,\varphi)_{ret}\right] \rightarrow \tilde{x}\left[z_{ret},(\theta,\varphi)_{dat}^{i,j}\right] \tag{12}$$

A two dimensional cubic spline interpolation routine has been implemented for this step of the forward function. The net result of this first step is altitude profiles on the z_{ret} grid at every measurement location.

The second step involved in the forward function is to degrade the profiles to the inherent vertical resolution of each instrument. This is a key point in reproducing real satellite data since different instruments and measurement techniques can have significantly different vertical resolution and this must be correctly taken into account. This step is accomplished by convolution of the profiles obtained in step one with the appropriate vertical averaging kernel for each instrument:

$$\tilde{x}\left[z_{ret},(\theta,\varphi)_{dat}\right] \rightarrow y^{i,j}\left[(z,\theta,\varphi)_{dat}\right] = \int_o^\infty dz_{ret}\; A^{i,j}(z_{dat},z_{ret})\,\tilde{x}\left[z_{ret},(\theta,\varphi)_{dat}\right] \tag{13}$$

The averaging kernels are taken to be triangle functions in altitude space. This simple analytical form effectively mimics realistic instrument averaging kernels. The effective width, defined by the FWHM of this triangle, is read in from the data files for each

instrument and is allowed to vary with altitude as it typically does for real instruments. Note that it is very easy in this formulation to include nadir viewing instruments which measure total integrated column amounts of trace gases, such as NASA's Total Ozone Mapping Spectrometer (TOMS) instrument, which makes global measurements of the ozone total column. By simply setting the averaging kernel $A^{i,j}(z_{dat}, z_{ret})$ equal to one we automatically obtain the integrated vertical column density. In this case $y^{i,j}$ has no altitude dependence and the measurement consists of only one value at each horizontal grid point.

Note that the forward model for the data assimilation problem, as implemented here, is nonlinear. This is because we have chosen a nonlinear horizontal interpolation scheme, the cubic spline, for the first step denoted in Equation (12). A simple linear interpolator could be used instead, resulting in a linear problem, but we felt that the spline interpolation routine provides better results. This issued will be studied in more detail in Phase II where a primary objective will be to push the algorithms to incorporate much larger grid spaces. In this case the additional time saved by using the one step linear model might be a crucial factor. In the current algorithm then the OPT routine uses either equation (6) or (7) to calculate the retrieved distribution depending on the relative dimensions of the data and retrieval spaces. In the typical case we have $NY > NX$ and OPT will utilize Equation (7) in the iteration.

4. Simulations.

4.1 Description of Simulations

In order to test the prototype assimilation algorithms and demonstrate the technical feasibility of the proposed approach, we have performed simulated retrievals. The specific problem addressed in the simulations was the retrieval of ozone distributions by assimilating data from two different types of ozone vertical profilers. We have chosen as the sample data sets three different instrument/platform combinations, which allow us to test the concept of assimilating data from data sets having different spatial sampling and vertical resolutions.

The simulations are performed as follows. First simulated data sets must be generated for each of the instruments. An independent forward model code has been developed specifically for this purpose. This code first reads in an external file containing measurement ephemeris for each instrument. These measurement locations, which are the $(\theta, \varphi)^{i,j}_{dat}$ grid points, correspond to the time period specified for the simulation and are generated using CPI's orbit propagation and measurement prediction models for each specific satellite/instrument configuration. The forward model code then generates simulated measured ozone profiles at all locations for all instruments as discussed in Section 3.2. The background ozone distribution used in the simulation, which defines the 'true' atmospheric state vector x, is taken from standard global ozone

climatology. This climatology consists of zonal (i.e., longitudinally) averaged ozone mixing ratios in 10-degree latitude bins for each month of the year. It is therefore only a two dimensional distribution and in order to add a longitudinal dependence we have introduced a simple longitudinal wave pattern into the distribution as follows:

$$x(z,\theta,\varphi) = x(z,\theta)\left[1 + \gamma \sin\left(\frac{k\varphi}{2\pi}\right)\right] \tag{14}$$

This produces a distribution with a longitudinal wave number k pattern of amplitude γ. Using a given ozone distribution in the forward model the simulated data sets are generated. These are then read in as data input to the assimilation models. For a realistic test it is important that the a priori, or first guess, ozone distribution used in the retrievals be significantly different from the true distribution used to generate the data. This is accomplished in practice by taking a different seasonal profile from the climatology as well as using different values for the wave pattern and amplitude in equation (14). The retrieved ozone distribution is then compared to the distribution used in the forward model to determine how well the assimilation routines have reproduced the true atmospheric state.

4.2 Results of OOAM/POAM III Simulation

The results of two different simulation scenarios will be presented here. In the first simulation we combine data from Naval Research Laboratory's (NRL) OAM series of instruments. These are solar occultation instruments, which are characterized by very high vertical resolution but relatively coarse spatial sampling. Two instruments were considered. The Polar Ozone and Aerosol Measurement (POAM) instrument is in a polar orbit on the SPOT satellite and therefore obtains measurements only at relatively high latitudes in each hemisphere. The Orbiting Ozone and Aerosol Monitor (OOAM) is essentially the same instrument but flies in a much lower inclination orbit (approximately 40 degrees) and therefore predominantly samples the mid-latitudes. These two instrument configurations are therefore highly complimentary in terms of measurement characteristics and global coverage and the problem of assimilating the two data sets is an important scientific objective for NRL.

Figure (1) shows a plot of the seasonal variation of measurement latitudes for the two instruments. Because of the nature of the solar occultation measurement, each instrument makes approximately 14 measurements per day around a circle of fixed latitude. The measurements are therefore separated by about 25 degrees of longitude. We have chosen a one-month time bin centered in July for the simulations, since during this month the two instruments overlap in the Northern Hemisphere. The top panel of Figure 1 shows the location of all OOAM and POAM III measurements in the month of July. Note that only POAM sunrise events occur in the Northern Hemisphere, whereas OOAM makes both sunrise and sunset measurements at these latitudes.

A significant problem to be addressed in the Phase II effort is how to extend the method to handle global-scale assimilation grids. Currently hardware limitations, especially memory requirements, limit the retrieval grid to relatively small sizes. However we are convinced that this is a limitation which will be overcome eventually and that a successful demonstration of the method on a grid of limited geographical size is a valid demonstration of the technical feasibility of the method. There is nothing inherent in the method that makes it less accurate as the number of grid points increases. It is simply a matter of expanding CPU time and memory requirements. For the OOAM/POAM III simulation we have chosen a retrieval grid which extends from 10 to 60 degrees in both latitude and longitude, in 10 degree bins, and from 10 to 40 km, in 2 km bins, in altitude. The total number of points in the assimilation grid, NX , is therefore 576. The bottom panel of Figure 1 shows the combined OOAM/POAM III measurement points included in the simulation. Note that the POAM III data points are concentrated along the upper edges of the grid space, as these are the lowest latitudes sampled by the instrument.

The OOAM and POAM III instruments are both characterized by FWHM vertical resolution of 1 km. For both instruments we have assumed random measurement errors of 5 - 10 % in the ozone profiles. The data altitude grids used for the simulation extend from 10 to 40 km with 1 km sampling for both instruments, and we have made the simplifying assumption of a uniform, unchanging grid for all measurement profiles. The data space dimension NY for this problem is therefore 992, significantly larger than the retrieval space dimension. The ozone distribution used in the forward model data simulation is obtained from the January climatological profile with a wave number 1 of amplitude 0.2 in the longitudinal variation. For the a priori we have used the July profile with wave number 2 of amplitude 0.3. In these simulations no measurement noise was simulated in the data – the effects of realistic measurement error will be included in Phase II.

In Figure 3 we have plotted the retrieval results for this simulation. Each panel represents one latitude/longitude point in the assimilation grid and has three profiles overplotted. The true ozone mixing ratio profile is plotted with a dashed line, the a priori profile is the dotted line, and the solid line represents the retrieved profile. In Figure 4 the same results are plotted in the form of relative errors, that is, difference profiles between the retrieved and true ozone mixing ratio at each grid point (solid lines). For reference the difference between the a priori and true profiles are also plotted in the dotted lines. This curve represents the starting point of the iterative retrieval. These results show that for the most part the algorithm is able to reproduce the true atmospheric state with high accuracy, even when starting from an initial distribution which is significantly in error. The worst convergence tends to be along the edge grid points. This is to be expected because these grid points generally get information from fewer data points, and those are always biased in one direction, toward the middle of the grid. This is particularly true in this case at the lowest latitude points where Figure 2 clearly shows that we are in a very data sparse region and the retrieved distribution obviously remains heavily biased toward the a priori. Note that the results are much better along the high latitude edges of the grid, presumably due to the dense spacing and subsequent

high information content, coming from the POAM III measurements. There are also presumably end point effects in the horizontal interpolation routines, which could tend to degrade the retrieval results at the edges of the grid. This issue will be investigated in more detail in Phase II.

4.3 Results of OOAM/MAS Simulation

For the second simulation we again choose to combine two ozone limb profilers. One of the instruments is again taken to be the NRL OOAM solar occultation instrument. For the second data set we have chosen the Millimeter Wave Atmospheric Sounder (MAS) instrument. MAS measures ozone by detection of millimeter wave emission features in the ozone spectrum. As an emission measurement the MAS is able to make measurements on the limb essentially continuously. It therefore has a much denser spatial sampling than the occultation data sets. However, the vertical resolution is significantly coarser, with an averaging kernel FWHM of 3 km as compared to 1 km for OOAM.

The MAS instrument flew three times, as part of the ATLAS payload on the space shuttle. Figure (5) shows the measurement locations, after the data have been binned in approximately two-minute bins, for the ATLAS 1 mission. For the simulation we have taken the ATLAS 1 time frame and combined the MAS data with the predicted OOAM measurements over the same time frame. For this case the most data-rich regions are the tropics and therefore we have chosen an assimilation grid which is similar in size to the one used in the first simulation but symmetric about the equator. It extends from –30 to 30 degrees in latitude and 0 to 60 degrees in longitude, in 10 degree steps. The altitude grid is the same as previously, 10 to 40 km in 2 km steps. The assimilation grid size, NX, for this case is then 784. Figure (6) shows in the upper panel the global OOAM and MAS data set for the time frame of the assimilation, and in the lower panel the data points which fall within the assimilation grid and thus were used in the simulation. The dimension of the data space vector, NY, for this case was 1056. The true and a priori ozone distributions used in the simulation were identical to the first simulation described.

Figures (7-8) show the retrieval results for the OOAM/MAS simulation. These plots contain the same information as Figures (3-4) for the OOAM/POAM III assimilation. Concentrating on Figure (8) we again see excellent results in retrieving the ozone distribution from the combination of these two data sets. The only grid points which show noticeable retrieval errors are again in data-sparse regions in the very corners of the grid. Again, the results of this simulation provide good evidence of the accuracy of the data assimilation algorithms.

5. Summary.

We have presented a detailed description of algorithms developed for the assimilation of multiple satellite remote sensing data sets into instrument-independent three-dimensional gridded distributions. As proposed, the assimilation problem has been formulated and solved as a general nonlinear retrieval problem, using the theory of optimal estimation. Detailed descriptions of the method, and the specific structure of the algorithms that result from its application to the problem of data assimilation, have been provided. The prototype algorithms that have been developed are already quite general in nature, easily incorporating an arbitrary number of independent data sets, with realistic treatment of satellite ephemeris and instrument characteristics. They may be applied to the assimilation of remote sensing measurements of any geophysical parameter of interest.

These algorithms have been tested for the specific case of assimilating global ozone distributions, by using simulated satellite data sets from satellite limb-viewing ozone measurements. Two scenarios have been studied in which data sets with significantly different spatial sampling patterns and vertical resolution have been assimilated to produce an ozone distribution. The three instruments considered in these simulations included the NRL OOAM and POAM III solar occultation instruments and the MAS millimeter wave emission measurement.

The results of these simulations clearly demonstrate the technical feasibility of the proposed approach. Starting from significantly different initial conditions, the retrieval algorithms were able to reproduce the correct ozone distribution from the simulated data with very small errors in both cases. The potential applications of a general, rigorous data assimilation algorithm are widespread because of the increasing dependence on, and sophistication of, satellite remote sensing data in both the defense and civilian sectors. Examples include the suite of polar orbiting satellites operated by DMSP and NOAA which provide climatological data for operational weather prediction, multi-platform scientific missions such as NASA's planned EOS program, and commercial earth remote sensing programs such as LANDSAT and the French SPOT program.

6. Figure Captions.

1. Seasonal variation of OOAM and POAM III measurement latitudes. The POAM III orbit is exactly periodic year to year whereas the OOAM orbit will precess with time.

2. The measurement locations for OOAM and POAM III during the month of July. The top panel shows the entire global coverage of each instrument for this month while the bottom panel limits the geographical area to that encompassed by the retrieval grid used in the simulation. The legend in the top panel plot explains the meaning of the different symbol types.

3. Retrieval results for the OOAM/POAM III simulation. Each panel shows profiles at one latitude/longitude point in the assimilation grid. Solid lines are the retrieved ozone mixing ratio profile, dotted lines are the a priori profile used in the retrieval, and dashed lines are the true profile used in the forward model simulation.

4. Retrieval errors for the OOAM/POAM III simulation. Solid line is the percent error in the retrieved profile at each latitude/longitude grid point. For comparison the dashed line is the error in the a priori profile, representing the starting point of the retrievals.

5. The MAS ATLAS 1 measurement locations after the data have been binned in 2-minute time bins. Open and solid circles denote daytime and nighttime measurements, respectively.

6. The measurement locations for OOAM and MAS during the ATLAS 1 mission. The top panel shows the entire global coverage of each instrument for this month while the bottom panel limits the geographical area to that encompassed by the retrieval grid used in the simulation. The legend in the top panel plot explains the meaning of the different symbol types.

7. Same as Figure (3) except for the OOAM/MAS data assimilation.

8. Same as Figure (4) except for the OOAM/MAS data assimilation.

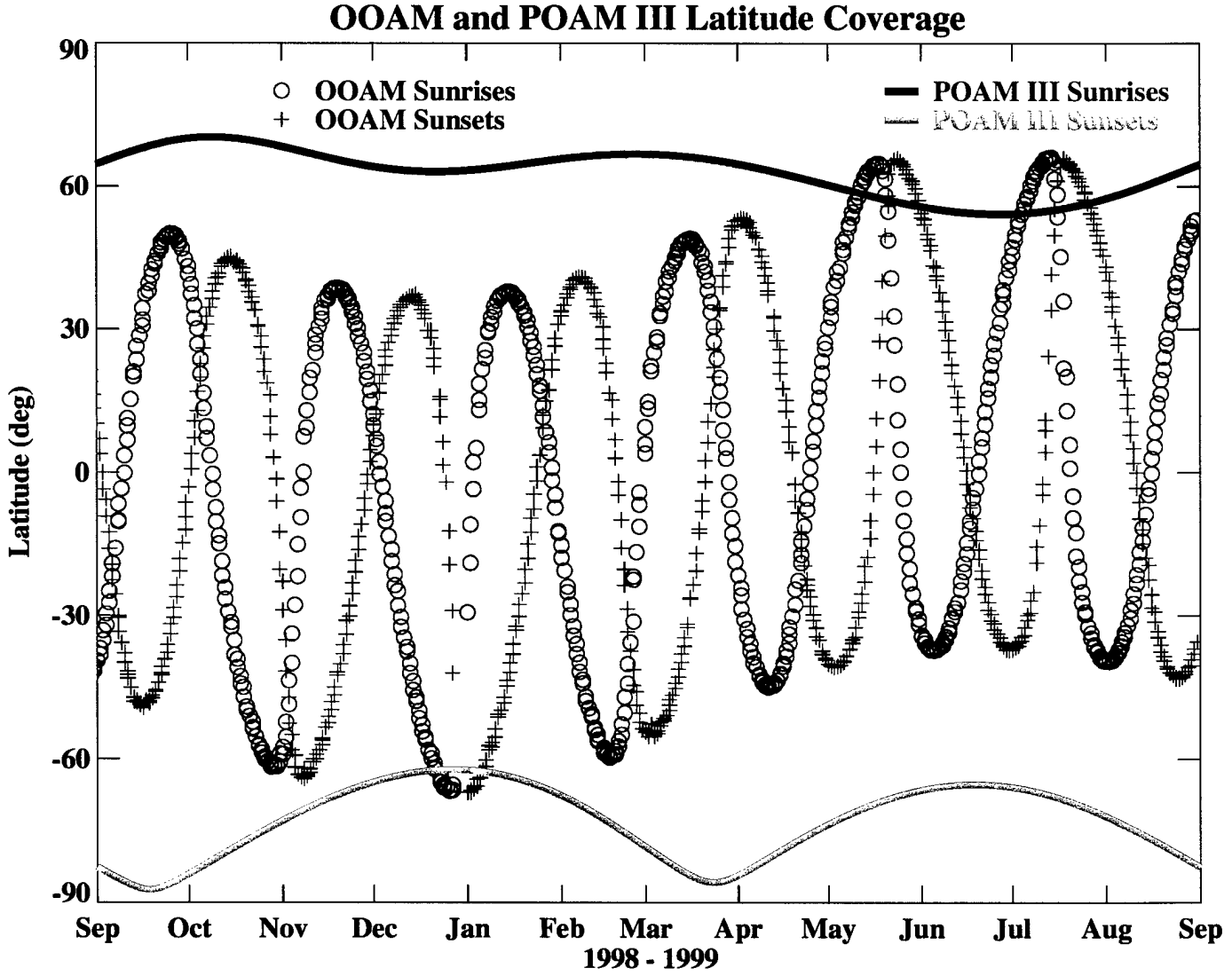

OOAM and POAM III Latitude Coverage
OOAM Sunrises
OOAM Sunsets
POAM III Sunrises
POAM III Sunsets
Latitude (deg)
90
60
30
0
-30
-60
-90
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
1998 - 1999

Figure 1

OOAM/POAM III Measurements

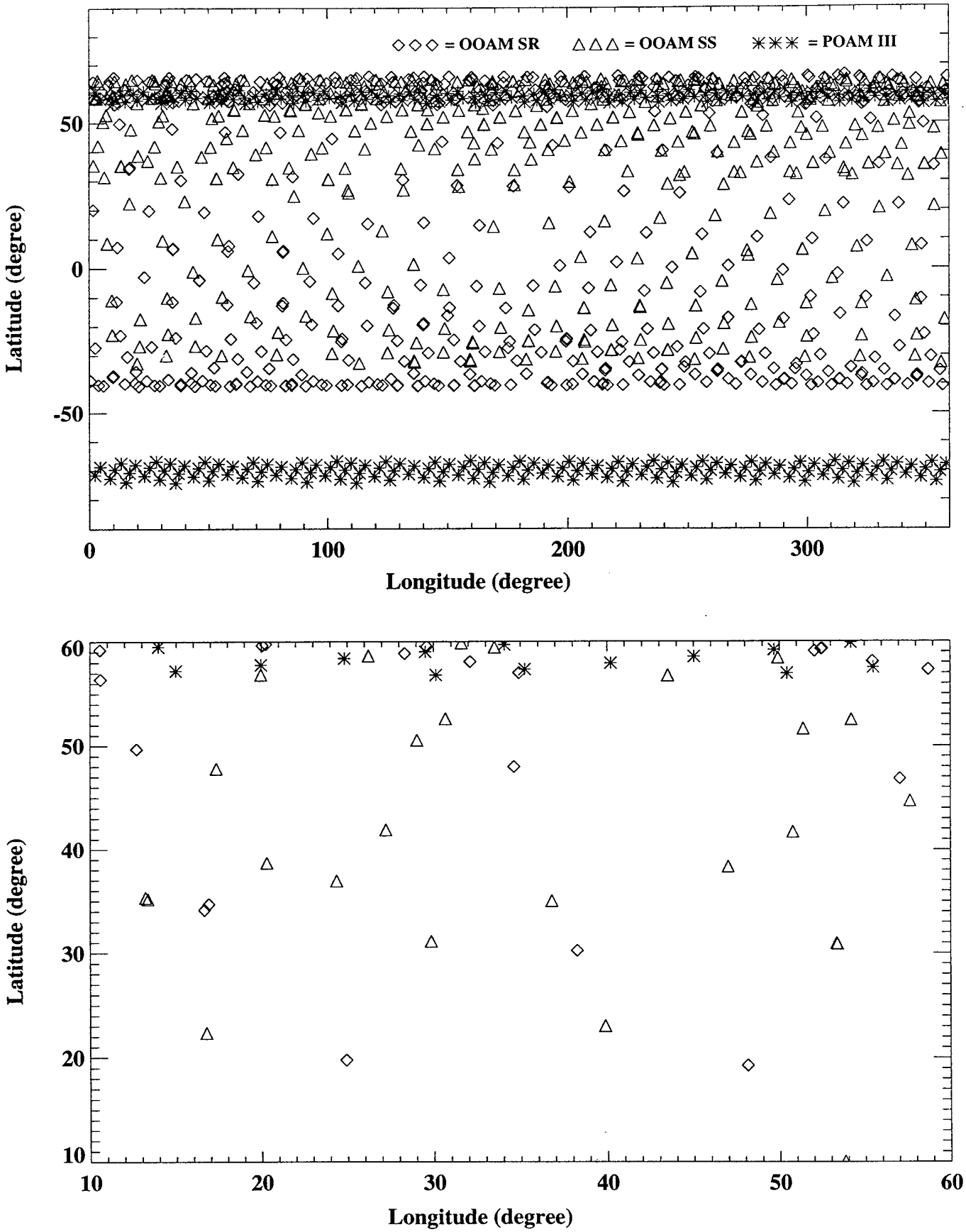
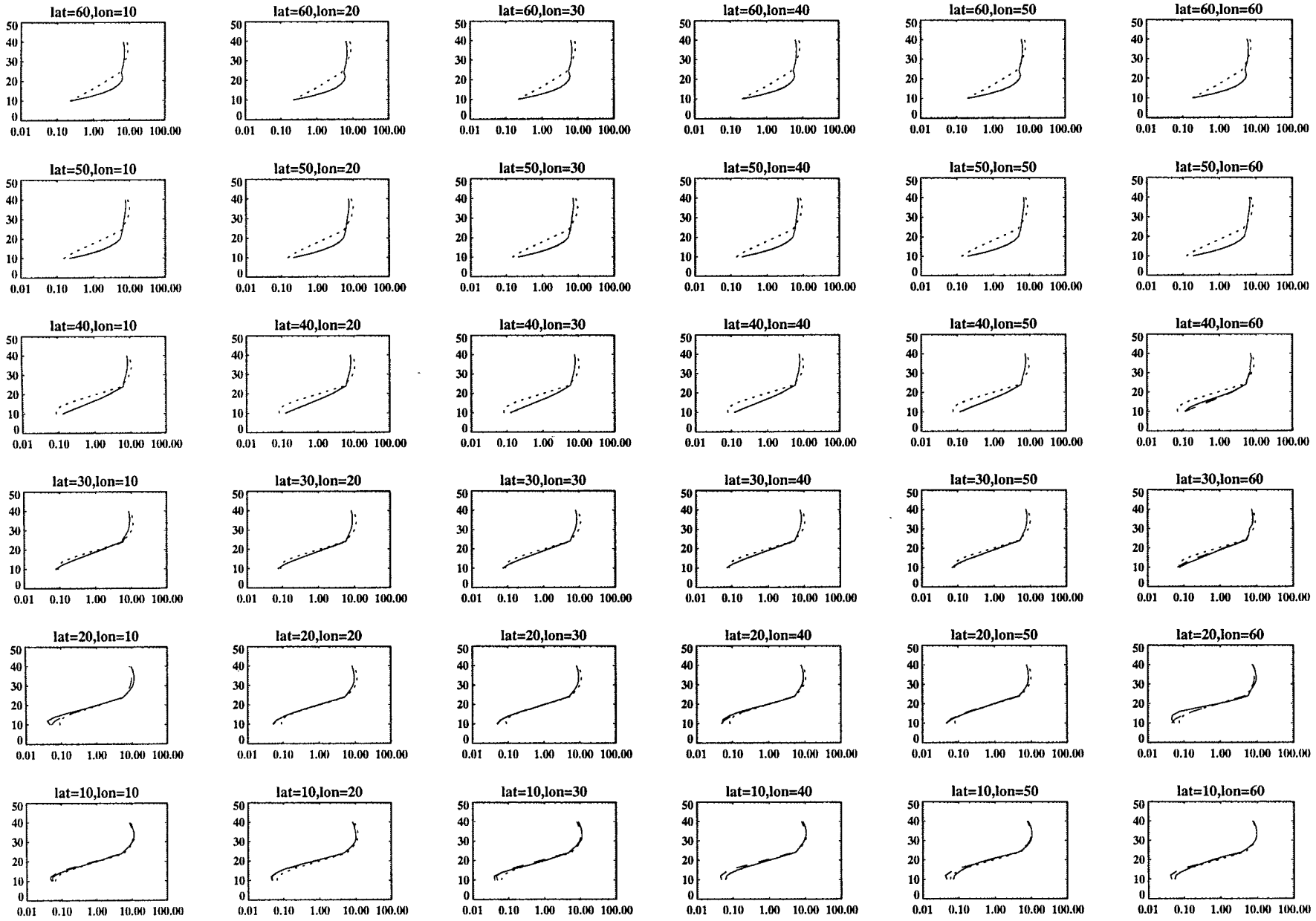


Figure 2

OOAM/POAM III Ozone Assimilation



Altitude (km)

Ozone Mixing Ratio (ppmv)

Figure 3

OOAM/POAM III Ozone Assimilation

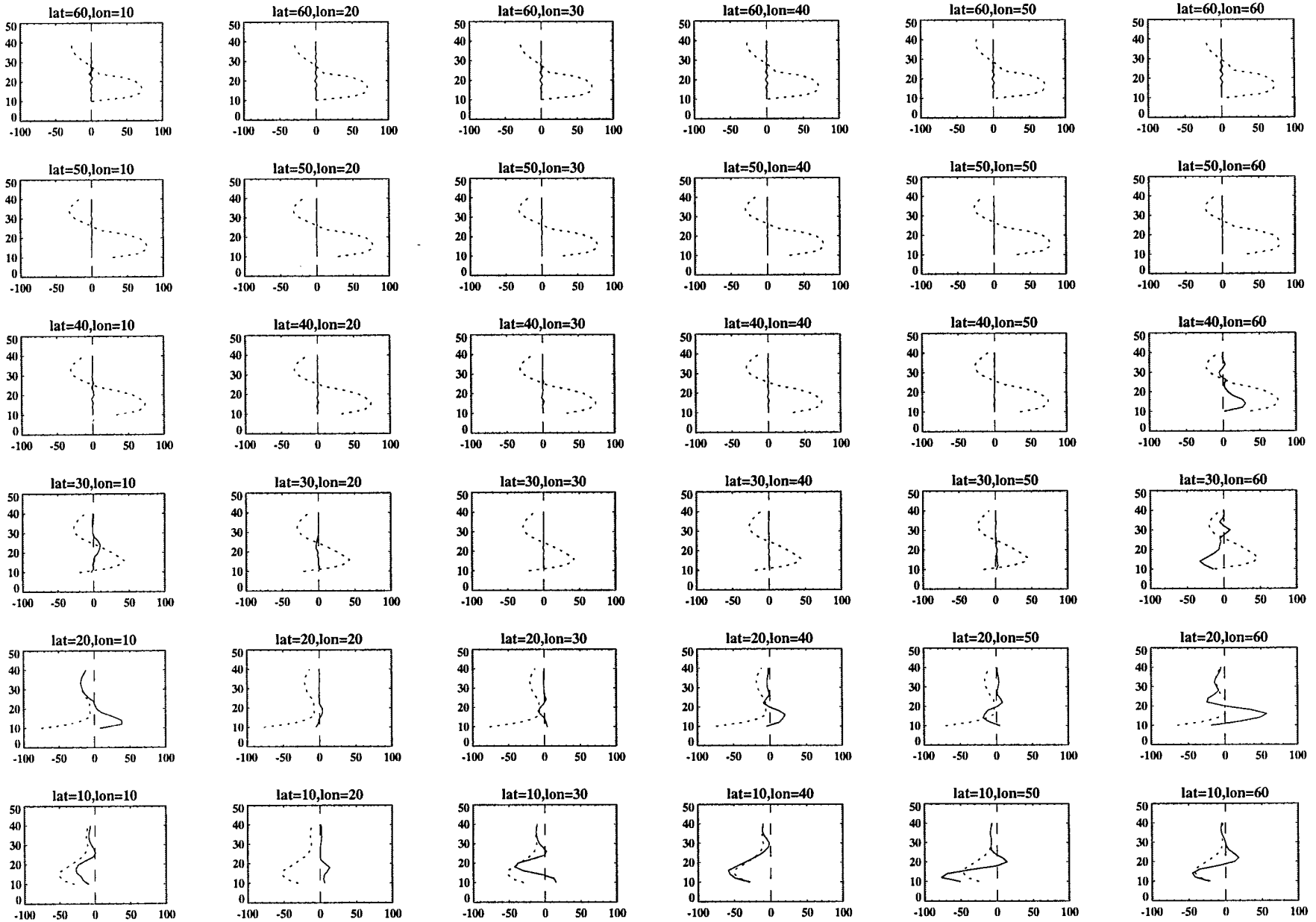


Figure 4

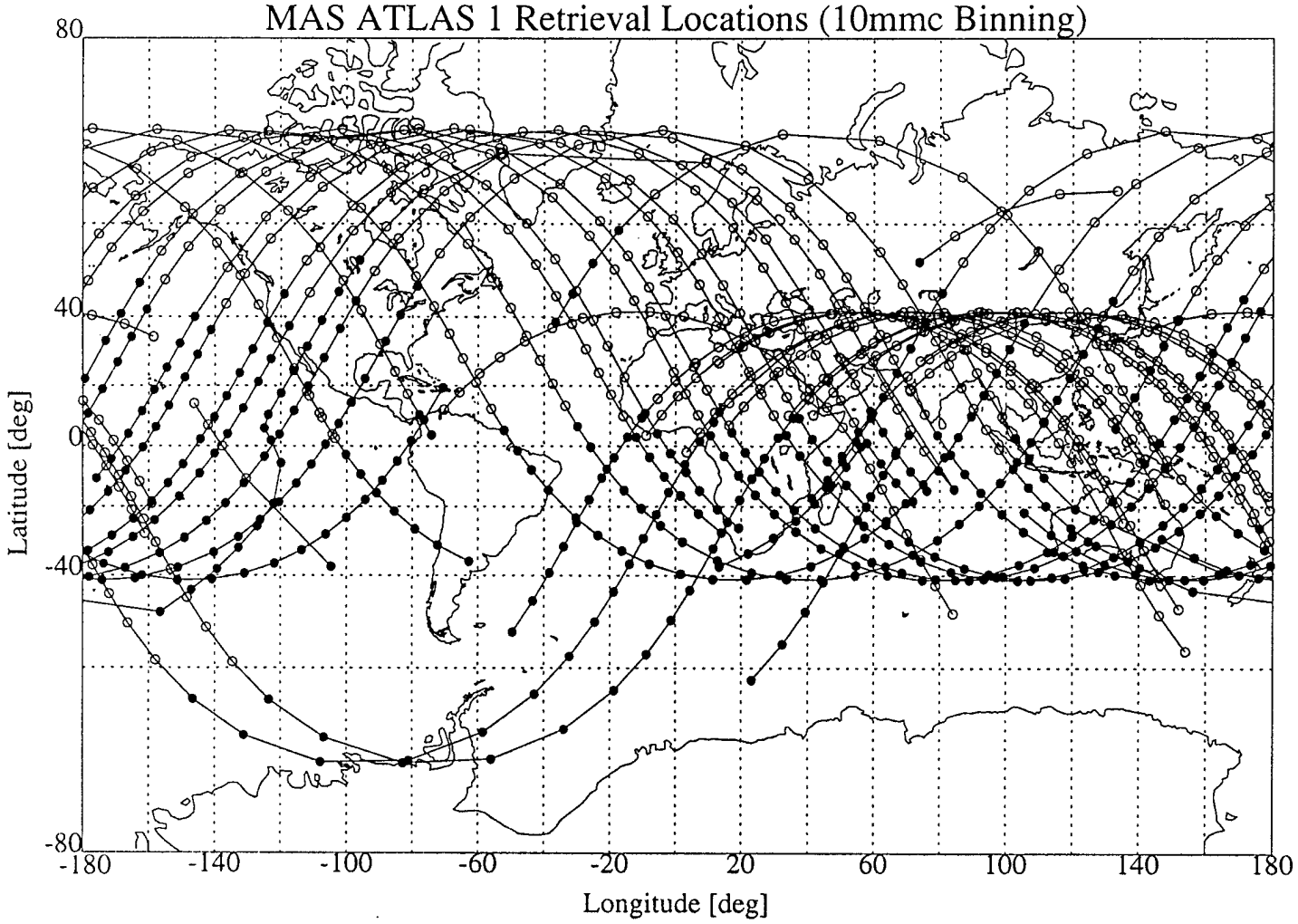

MAS ATLAS 1 Retrieval Locations (10mmc Binning)
80
40
0
-40
-80
Latitude [deg]
-180
-140
-100
-60
-20
20
60
100
140
180
Longitude [deg]

Figure 5

OOAM/MAS Measurements

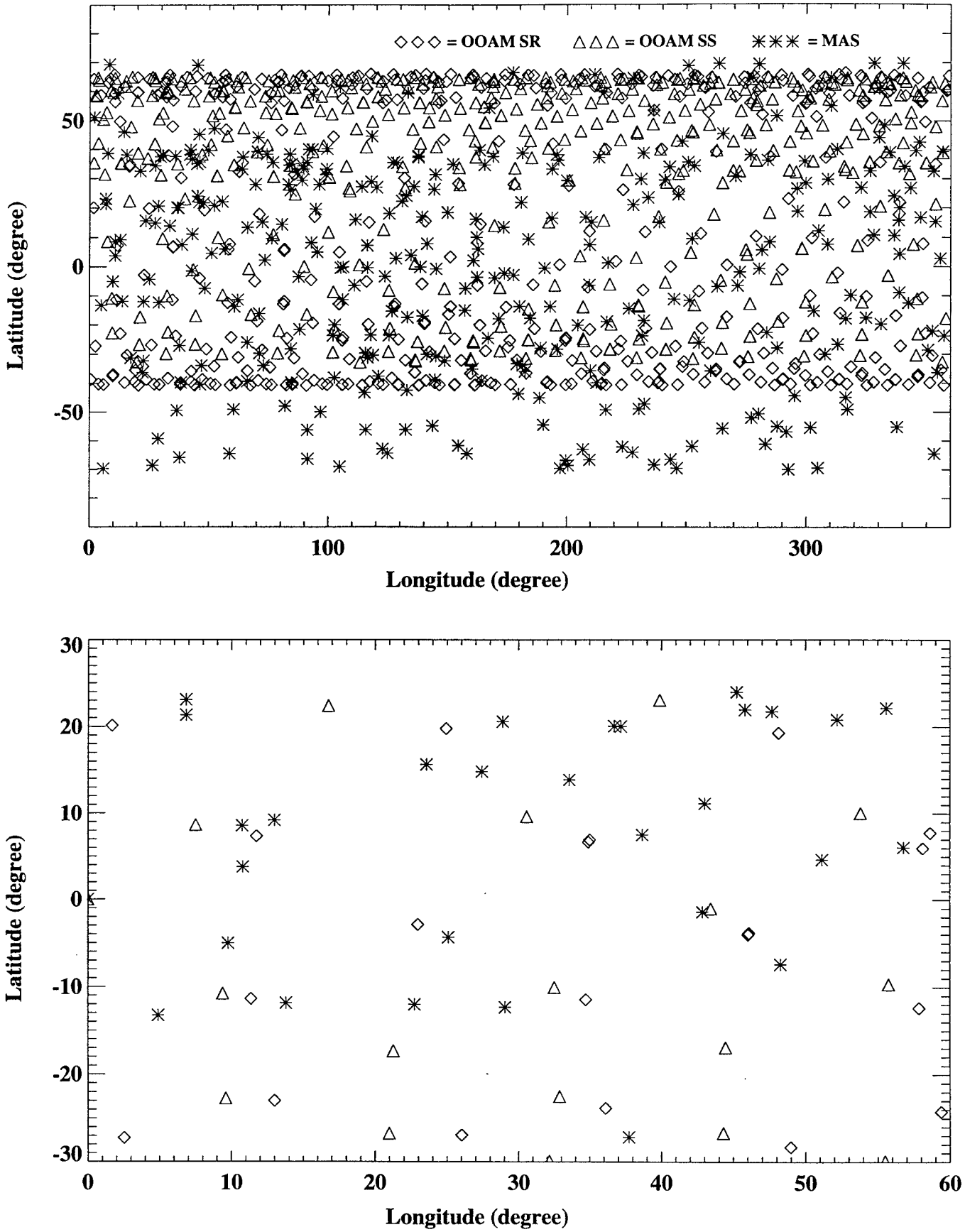
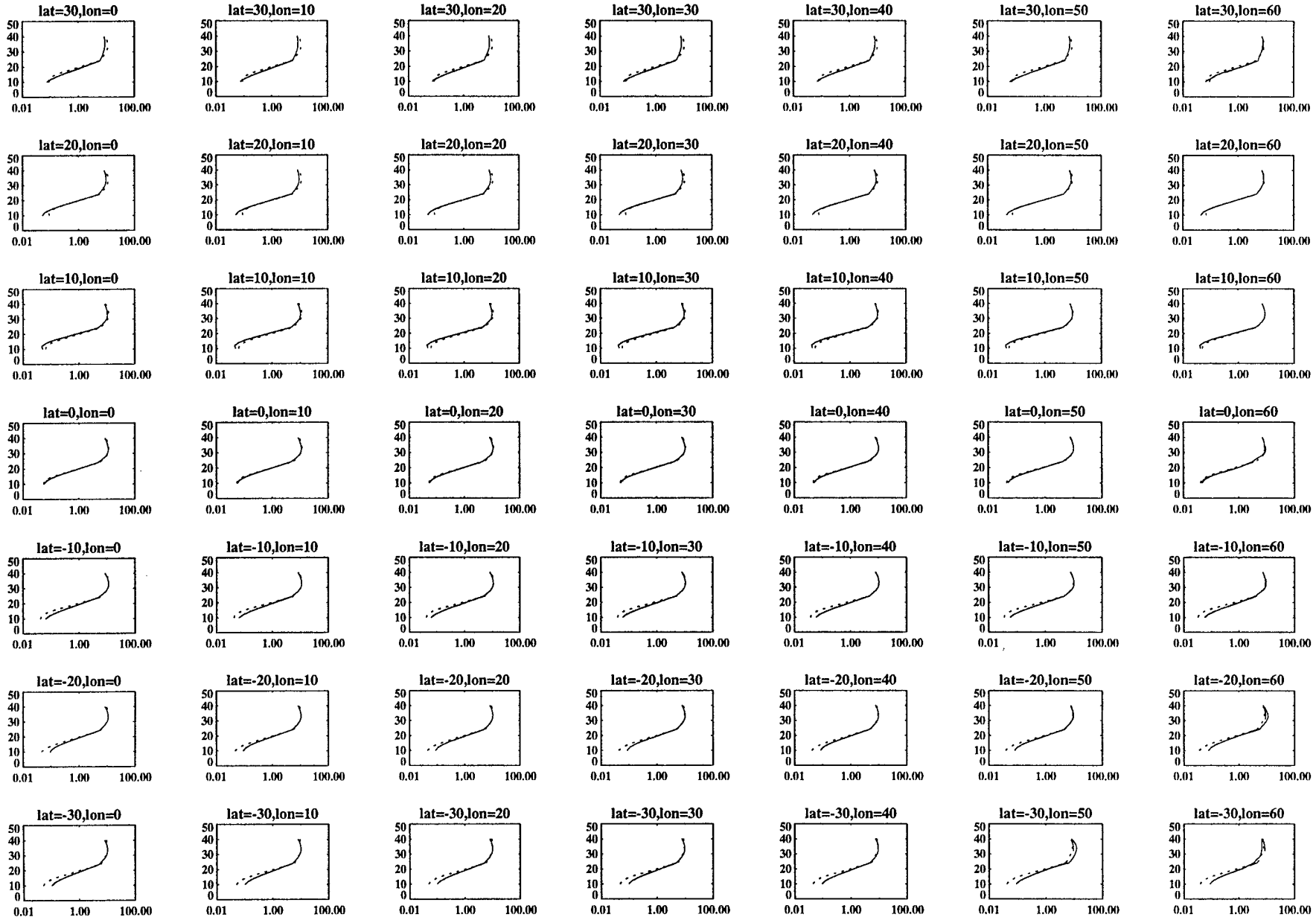


Figure 6

OOAM/MAS Ozone Assimilation



Altitude (km)

Ozone Mixing Ratio (ppmv)

Figure 7

OOAM/MAS Ozone Assimilation

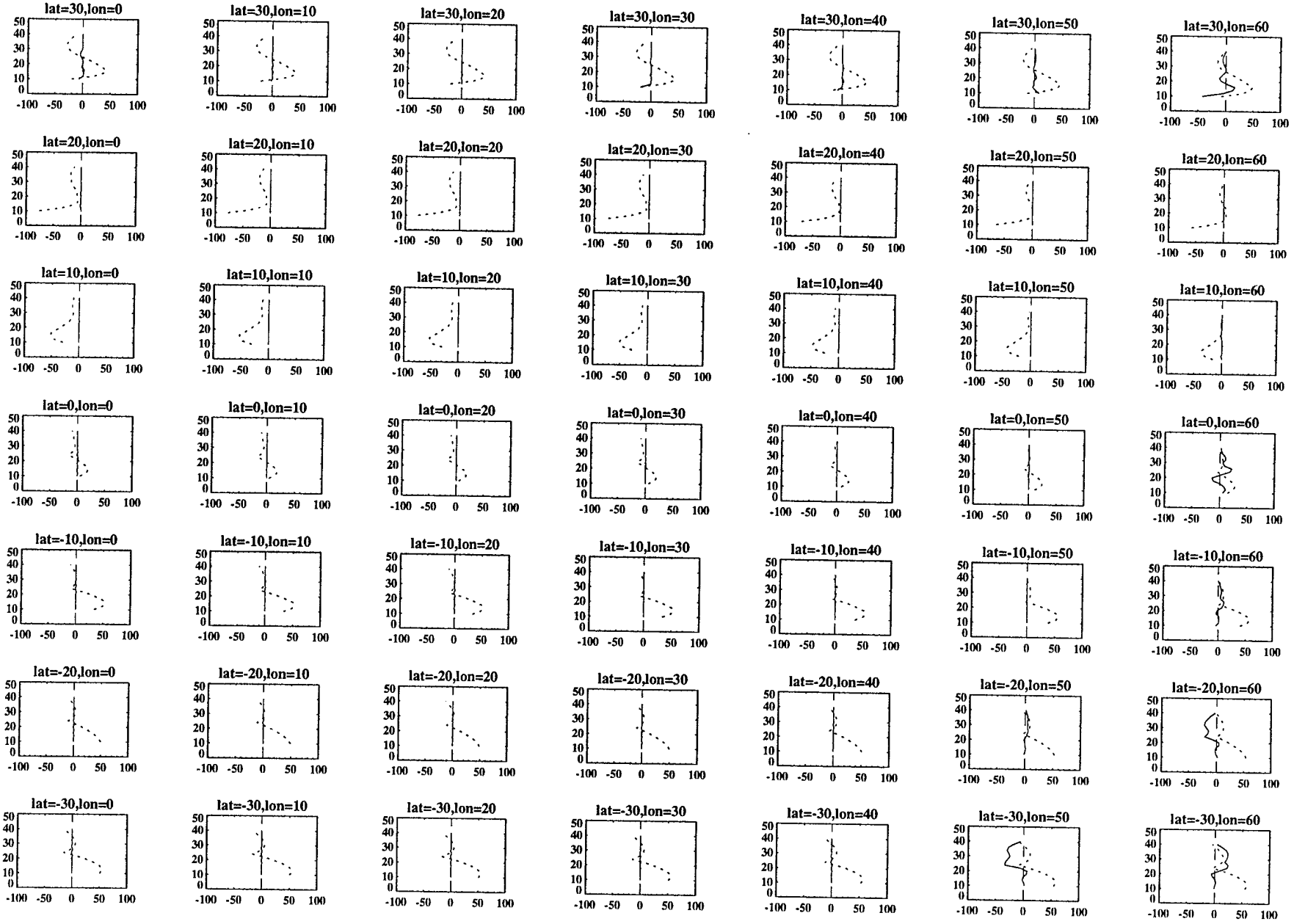


Figure 8