Sampling uncertainties in observing the global aerosol with a nadir-viewing satellite lidar

Santa Maria, M.*[1, 2], Winker, D.[2]

[1]Analytical Services and Materials Inc., [2]NASA Langley Research Center

ABSTRACT

In an attempt to better understanding climate and better comprehend the effects of clouds and aerosols on the Earth's Radiation Budget, NASA has been developing several satellite missions. Among them, the Cloud-Aerosol-Lidar and Infrared Pathfinder Spaceborne Observation (CALIPSO) mission will observe clouds and aerosols with a combination of lidar and passive instruments. CALIPSO will fly in formation with EOS Aqua, EOS Aura, Cloudsat and Parasol. This novel satellite formation will provide a unique comprehensive data set of cloud and aerosol optical and physical properties, and radiative fluxes.

In this paper, the characterization of global aerosol properties with sparsely sampled observations is investigated using a dataset of aerosol optical depth (AOD) from the MATCH climate model. MATCH is an offline Chemistry and Transport Model (ChTM) primarily developed by NCAR that includes a number of aerosol sources as well as a variety of transformation and removal mechanisms. The CALIPSO satellite is "flown" through this dataset and the aerosol optical depths at the CALIPSO footprint locations are sampled to produce an AOD subset. Averages computed from the subset are compared with averages from the full model output to investigate the magnitude of uncertainties due to sparse sampling of the aerosol field.

Initially, uncertainties in satellite sparsely sampled measurements of global aerosol distribution are quantified in terms of zonal averages. The goal of this effort is to determine the correct satellite average scaling to accurately represent global aerosol coverage. Ultimately, sampling errors will also be assessed at regional scales.

Keywords : Lidar, aerosols, climatology, sampling, CALIPSO, MATCH, chemical and transport model.

1. INTRODUCTION

In an attempt to better understanding climate, and reduce the uncertainties in the effects of clouds and aerosols on the Earth's Radiation Budget, NASA has been developing several Earth System Science Pathfinder (ESSP) research missions. Among them, the Cloud-Aerosol-Lidar and Infrared Pathfinder Spaceborne Observation (CALIPSO) mission will address these uncertainties through a combination of lidar and selective passive imager measurements (1). CALIPSO has been scheduled for launch in 2005, to operate as part of the A-train satellite constellation containing EOS Aqua, EOS Aura, Cloudsat and Parasol. This novel satellite formation will provide for the first time a unique comprehensive data set of cloud and aerosol optical and physical properties, and radiative fluxes.

* Santa Maria, M. NASA Langley Research Center, Atmospheric Sciences Division, MS 435, Hampton, VA 23681;
ph. 757-864-7189; e-mail: m.santamaria@larc.nasa.gov

Lidar Remote Sensing for Industry and Environmental Monitoring V,
edited by Upendra N. Singh, Kohei Mizutani, Proceedings of SPIE Vol. 5653
(SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15 · doi: 10.1117/12.583098

For over 30 years, ChTMs have been developed in an attempt to reproduce and forecast the behavior of the atmosphere. ChTMs provide theoretical aerosol and trace species concentration profiles in a wide range of temporal and spatial scales, based on general circulation and chemical transport methods, source emission inventories and meteorological modules that simulate the precipitation rates and wind flows of the planet (2). Throughout years, ChTMs have incorporated chemical schemes of increasing complexity, have accounted for physical processes, and progressively included feedback effects between climate, meteorology and chemistry. The Model of Atmospheric Transport and Chemistry (MATCH) is an offline transport model developed primarily by the National Center of Atmospheric Research (NCAR) and the Center for Clouds, Chemistry and Climate (C^4) (3).

The simultaneous analysis and scrutiny of satellite mission data, such as from CALIPSO, and predicted aerosol distributions from third generation ChTMs, such as MATCH, represents an unprecedented opportunity to mutually benefit and improve our understanding of climate.

In this study, averaged optical thickness values corresponding to satellite overpasses accumulated over grid boxes, are compared against model-averaged diagnostics obtained from 6-hour samplings over the entire globe. Results show that differences between irregularly (satellite) and regularly (model) sampled diagnoses are the greatest in the Northern hemisphere where the number of located sources of aerosols is higher. Consequently, the variance of the AOD increases. Likewise, the variance decreases for longer time periods averaged.

2. GOALS OF THE STUDY

Aerosol measurements from CALIPSO, and from CALIPSO in combination with other A-train instruments will be used to derive improved estimates of aerosol radiative forcing of the climate system. CALIPSO observations are nadir viewing only, however, so the sampling of the global aerosol field is very sparse. There will be uncertainties in the global (or regional) mean aerosol forcing, which are due to this sparse sampling. We expect, however, that these sampling errors are unbiased. Therefore, space and time averaging can reduce the magnitude of the sampling errors. The question is then: What spatial and temporal averaging is necessary to reduce sampling errors to acceptable levels? Alternately, on what spatial and temporal scales does the sparse nadir-only sampling of CALIPSO give accurate estimates of global aerosol properties? The answer depends partly on the spatial and temporal correlation of the various relevant aerosol properties. When aerosol properties at two nearby locations are highly correlated there is no need to sample both because measurements at one location can be used to accurately predict the other. If they are uncorrelated, then the representation of both locations by a measurement at only one results in a sampling error. The higher the degree of correlation, the fewer the number of samples needed to accurately represent the global aerosol. The lower the degree of correlation, the more averaging is required to reduce the sampling noise to acceptable levels.

In this study we used AOD as a measure of our ability to characterize the global aerosol field. We take MATCH daily AOD dataset as truth, and investigate the ability of nadir-only satellite observations to represent the mean properties of the global aerosol at different space and time scales. The model dataset employed is part of the CERES/SARB project results, and consists on total column daily average of AOD per species and total (4). Sampling biases were estimated from intercomparisons of several time frames zonal and regional means computed from both the full MATCH output and from the subset along the CALIPSO ground track.

3. CORRELATION STUDY

An h-scatterplot is a convenient graphical tool to study the correlation between homogeneously spaced data. An h-scatterplot shows all possible pairs of data values whose locations are separated by a certain distance in a particular direction (5). For the spatial correlation study of the model climatology, the separation between two data points will be denoted by the vector h[x,y], meaning that we have taken each data location and paired it up with the data point located x longitude increments to the East and y latitude increments to the North.

The shape of an h-scatterplot cloud tells us how continuous the data values are over a certain distance in a particular direction. The closer the pairs plot to a x=y line, the more similar the data are. To quantify its degree of correlation, we used the correlation coefficient, ρ, and the correlation function, ρ(h). The correlation coefficient ranges from –1 to 1, being inversely proportional to the width of the cloud; that is to say, the wider the cloud gets, the closer to zero the correlation coefficient becomes.

$$\rho(h) = \frac{C(h)}{\sigma_{-h}\sigma_{+h}} \qquad (1)$$

where C(h) is the covariance function:

$$C(h) = \frac{1}{N} \sum_{(i,j)/h_{ij}=h} v_i v_j - m_{-h} m_{+h} \qquad (2)$$

where N is the number of data pairs whose location distances h, m_{+h} is the mean value of all the data points whose locations are +h away from some other data location, and namely m_{-h} is the mean value of all the data points whose locations are –h away from some other data location.

σ_{-h} is the standard deviation of all the data values whose locations are –h away from some other data location and σ_{+h} is the standard deviation of all the data values whose locations are +h away from some other data location.

$$\sigma_{-h}^2 = \frac{1}{N} \sum_{j/hij=h} v_i^2 - m_{-h}^2 \qquad (3)$$

Figure 1 shows the h-scatterplot of MATCH AOD for January 2001, for all the data pairs located at the same latitude and easting one grid longitude increment, that is to say for h[1,0]. In this case the correlation coefficient is 0.973, indicating a high level of data correlation along the latitude bands.

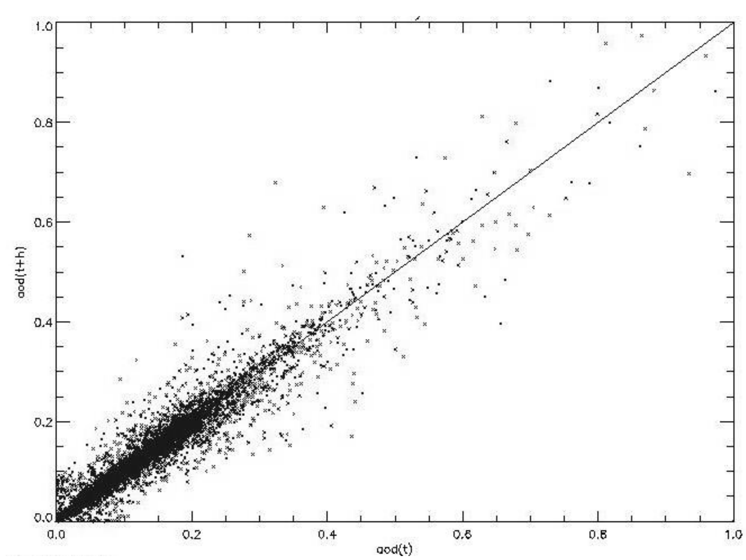

Figure 1. h-scatterplot for MATHC January, 2001 AOD

Figures 2, 3 and 4 picture the correlation function in the longitudinal, h[x, 0], latitudinal, h[0, y], and transversal, h[x, y], direction. In this case, for the transversal study, x equals y. A similar study has been done for every month of the year 2001, and the results yielded are consistent.

Because of the dominant planetary winds (jet stream), the aerosols are transported from their sources to the East. Consequently, as seen in Figure 2, the longitudinal correlation function is almost symmetrical, meaning that the aerosol concentration and composition on a certain latitude band is highly correlated both forward and backward from a certain location. However, the correlation of the data in a longitude band or transversally is very low. In Figures 3 and 4 the correlation function rapidly decreases as the separation between the data pairs increases. These results yield us to believe that the correct spatial averaging would correspond to a grid with wider longitudinal than latitudinal spacing.

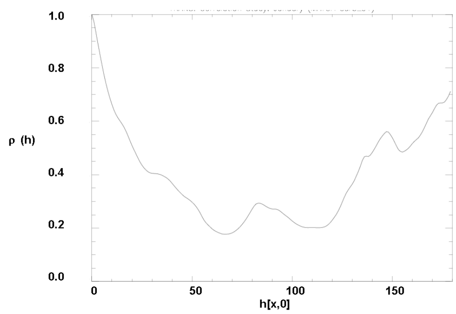

Fig 2. Longitudinal ρ(h) for h[x, 0]

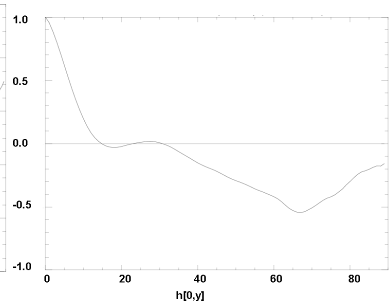

Fig 3. Latitudinal ρ(h) for h[0, y]

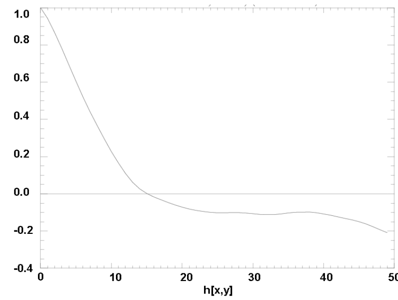

Fig 4. Transversal ρ(h) for h[x, y]

The geographical distribution of the model AOD time variance serves as a measure of the model's ability to capture individual aerosol events. A study like this, allows us to guesstimate the error introduced by time averaging of missing satellite observations or noise-contaminated profiles as a function of location. For specific quantification of errors, a finer model AOD time sampling is required. The time variance is calculated as average squared difference of the model data from their mean

$$\sigma^2 = \frac{1}{n}\sum_t (x_t - m)^2 \qquad (4)$$

where x_t is model AOD at certain location at a given time t; m is the mean of the data series and n is the total number of elements.

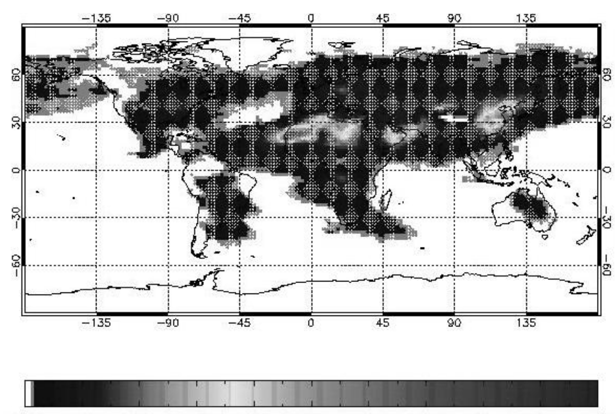

Figure 5. Time variance of MATCH AOD for year 2001

As seen in Figure 5, the highest model sensitivity to AOD seasonal variability corresponds to typical biomass burning areas such as the Amazon in Brazil; arid and deserted zones such as the Saharan desert with their dust plumes drifting over the Atlantic ocean; and industrialized regions such as Eastern Asia or North America. Because MATCH predetermines the maritime aerosol composition according to the Optical Properties of Aerosols and Clouds (OPAC) software, the temporal variability is incorrectly neglected over the oceans, emphasizing the deficient model simulation of sea salt and maritime aerosols. The low variability might also be influenced by low AOD values on that area and unvarying monthly mean wind speeds.

4. SAMPLING UNCERTANTIES

In this paper, CALIPSO orbit track is used to sample daily-average geographical distributions of MATCH optical parameters. Next, the model and the satellite latitude-band averaged AOD for different time frames; specifically, daily, weekly, monthly, seasonal and yearly, are compared.

CALIPSO will fly in 705-km circular sun-synchronous polar orbit. A full orbit cycle is completed every 16 days; that is to say, after 16 days, the lidar is pointing at the same initial ground spot and repeats the ground track of the previous 16 days. The lidar profiling frequency is once per second, corresponding to a spacing of 7 km along the track. The "nearest neighbor" AOD value for the corresponding time period is attributed to each of the lidar footprint locations. Nearest Neighbor is defined as MATCH physically closest grid point.

MATCH original grid, herein referred to as G1, consists of 192 bins of longitude and 94 bins of latitude, which corresponds to longitude increments of 1.875º with origin in the Greenwich meridian and ranging from [0º, 360ºE] and latitude increments of 1.9º with origin at the south pole parallel and ranging from [-90ºS, 90ºN]. In addition to G1, two other global grids are defined; G2 consisting of 36 longitude bins of 10º increments and 36 latitude bins of 5º increments and G3 consisting of 12 longitude bins of 30º increments and 12 latitude bins of 15º increments. Each of the G2 and G3 grid points is assigned the mean AOD value of the G1 neighboring grid points.

A structure with a number of entries equal to the number of satellite profiles of a full 16-day orbit cycle is created. The latitudinal and longitudinal coordinates of each satellite profile are read from the orbit file (Calipso_GT16_1s.report available from Kathy Powell[*]) and stored. Next, the indexes of MATCH G1, G2 and G3 nearest neighbor grid point latitude and longitude variables are attributed to each CALIPSO profile. The structure is now ready to be sampled for any period of time. Elapsed time is determined by number of records read; being one record equivalent to one second elapsed. When the number of records read equals the total number of records of a full rotation, the pointer for number of records read is reinitiated, so that the location indexes are back to be the initial point ones, while the time counter keeps adding seconds until the time unit to be averaged is completed, being this unit either one day, one week, one month, or three months (season). The pointer for the model time period corresponding to the profile being read is determined by dividing the total number of seconds elapsed by the number of seconds in one day (don't forget that the model data available are daily AOD means) and adding that to the initial model time period of the satellite orbit commencement.

The zonal mean study is obtained by dividing the accumulated sum of each satellite profile AOD over a latitude band for a determined period of time by the number of samplings in that band during that period of time. In the case of the model, the number of samplings equals the number of days contained in the total period of time under analysis times the number of longitude bins. The number of satellite samplings is generally higher than the number of model samplings over a latitude band.

The commencement time period is randomly selected using IDL's randomu(x) function where x is the variable seed used to initialized the uniformly distributed random number distribution. The starting point of the CALIPSO orbit is randomly selected as well. This way we guarantee an unbiased study of short-term time average sampling uncertainties, since the averaging of several short-time band averages usually averages out the differences between them and other longer time units studied.

5. RESULTS

The differences between the model and the satellite zonal mean averaging are compared in terms of percent difference, figures (b), and absolute difference, figures (c). The percent difference is defined as the ratio between the difference

[*] Kathy Powell. NASA Langley Research Center, Hampton, VA, 23681. k.a.powell@larc.nasa.gov

between satellite zonal mean AOD and the true zonal mean AOD value, over the true value. The percent difference results can be misleading if the AOD values are low. Thus, the absolute error is also included in the studied. CALIPSO will not orbit directly over the poles. The minimum satellite latitude profiling corresponds to 8.19º and the maximum satellite latitude profiling is 171.8º. Hence there would be 4 southern and 4 northern G1 latitude bands, and 1 southern and 1 northern G2 latitude bands without satellite records. In those cases, both the percent difference and the absolute difference are set to zero. The satellite AOD average for those latitude bands is left undefined.

The first row of graphs on Figure 6 shows the AOD latitude band average results for G1, the second row corresponds to G2 and the third row to G3 results. The black solid curve corresponds to MATCH and the dotted-dashed blue line to CALIPSO. Logically, as the grid gets broader, as the spatial averaging increases, the error curves are smoother and milder.

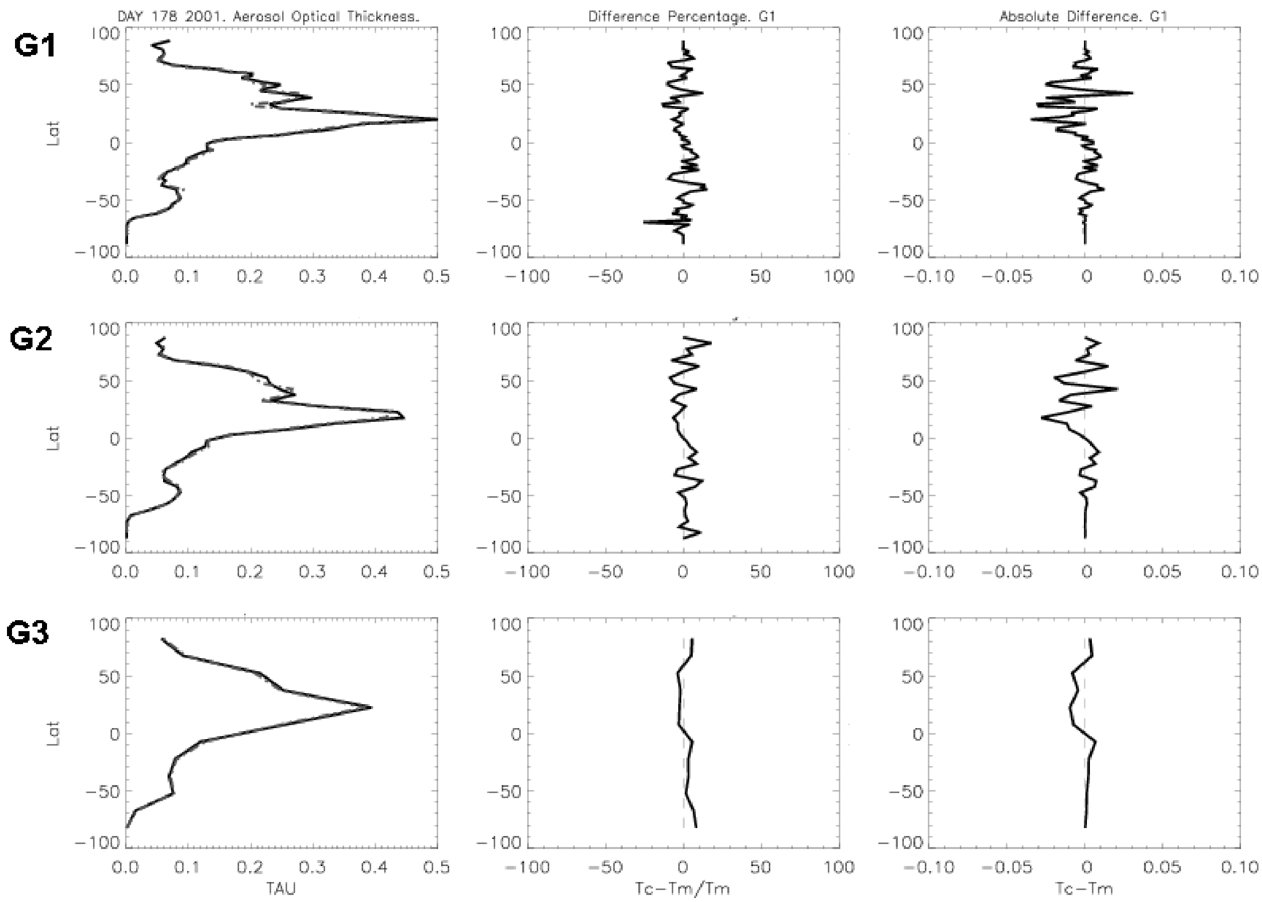

Figure 6. G1, G2, G3 daily zonal mean averaging comparison (mission day 178)

As a general rule, the largest disagreements between the model and the satellite AOD latitude band averages for any grid and for any time period studied are in the Northern hemisphere. This can be due to the fact that the northern hemisphere comprehends a higher concentration of industrialized areas, deriving in more localized aerosol sources and higher AOD gradients.

Figure 7 compares the absolute error of the sampling bias due to the irregular geographical distribution of satellite samplings at individual grid points for different time units. The dotted blue line represents the daily zonal means, the dashed yellow line the weekly zonal means, monthly zonal means are traced by the dashed and dotted orange line and the seasonal (autumn) zonal means are marked by the dashed double dotted red line. Each graph corresponds to a different spatial grid, from left to right, G1, G2 and G3. When the spatial grid is as ample as in the case of G3, there is not a significant error reduction by increasing the time unit averaged. These results would be corroborated later in the analysis of each grid and time unit absolute error variance by a rather flat G3 curve. Both in the case of G1 and G2 the error function is significantly smoother when the time unit averaged equals a week or longer.

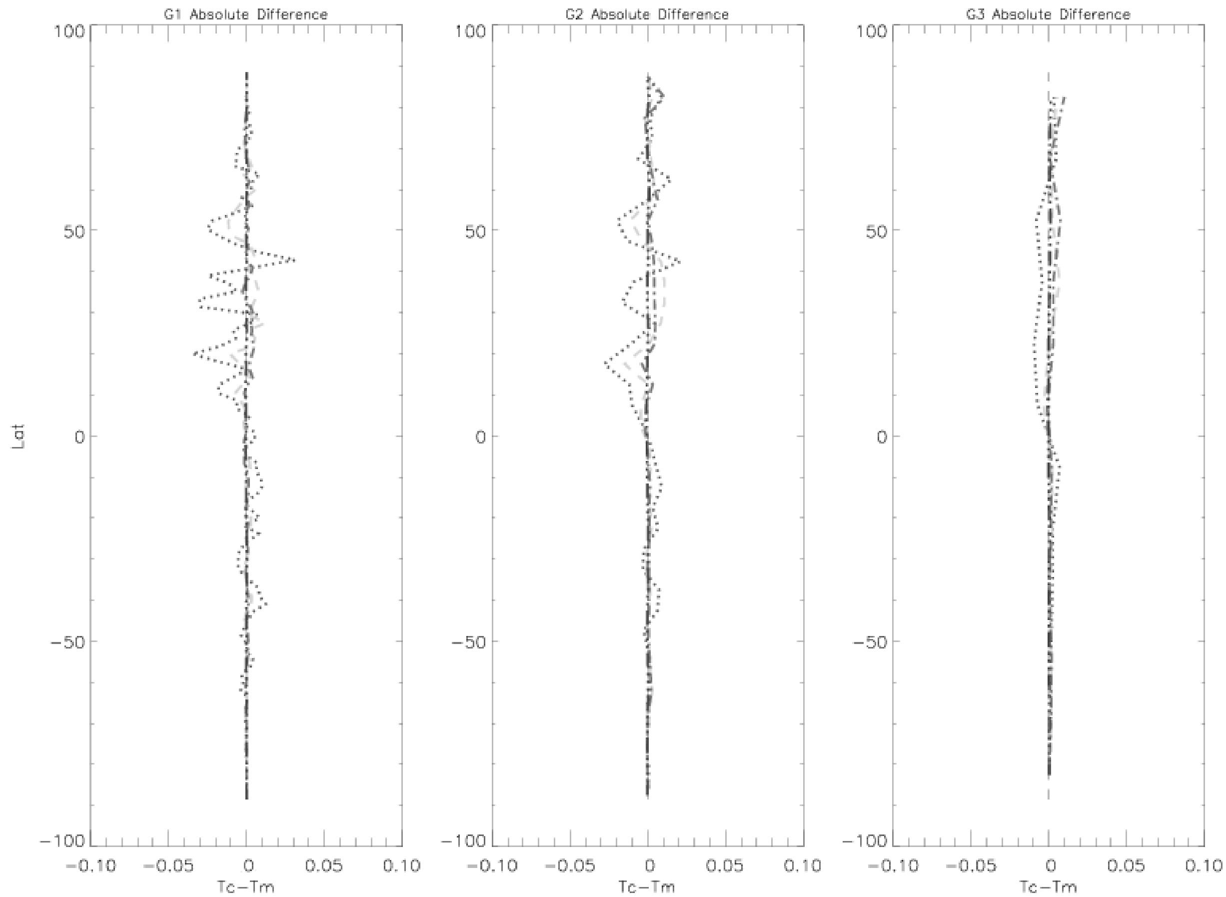

Figure 7. Absolute error of latitude band averaging for G1, G2 and G3 for several time units

Figure 8 shows the mean absolute error between the true AOD values and the satellite data and its standard deviation for different time units for each grid; the orange diamonds correspond to G1, the blue asterisks to G2 and the yellow squares to G3. The standard deviation is a measure of the spatial variability of the distribution. For instance, monthly averages have absolute errors of the same order of magnitude although slightly higher than weekly averages for each of the three grids proposed; however, the standard deviation of monthly averaging is almost zero whereas weekly averaging would still introduce some significant error peaks at certain areas in the case of G1 and drastically in the case of G2. As anticipated G3 absolute error variance is very low for any time averaging over a day.

It is interesting to notice the contrast of the error history between the Northern and the Southern hemisphere. In the case of the Southern hemisphere, the low absolute errors and the null standard deviation of the distribution indicate that only a mild time and spatial averaging is required.

In the case of seasonal averages, the data accounted for in these graphs corresponds to the fall season. In the case of each grid both the mean absolute error and the variance are negligible. Whereas the uncertainties of satellite sampling would be neutralized, such an ample averaging is only possible when analyzing long-term climatologies; otherwise sudden events, for instance natural aerosol occurrences such as a volcano, would be masked.

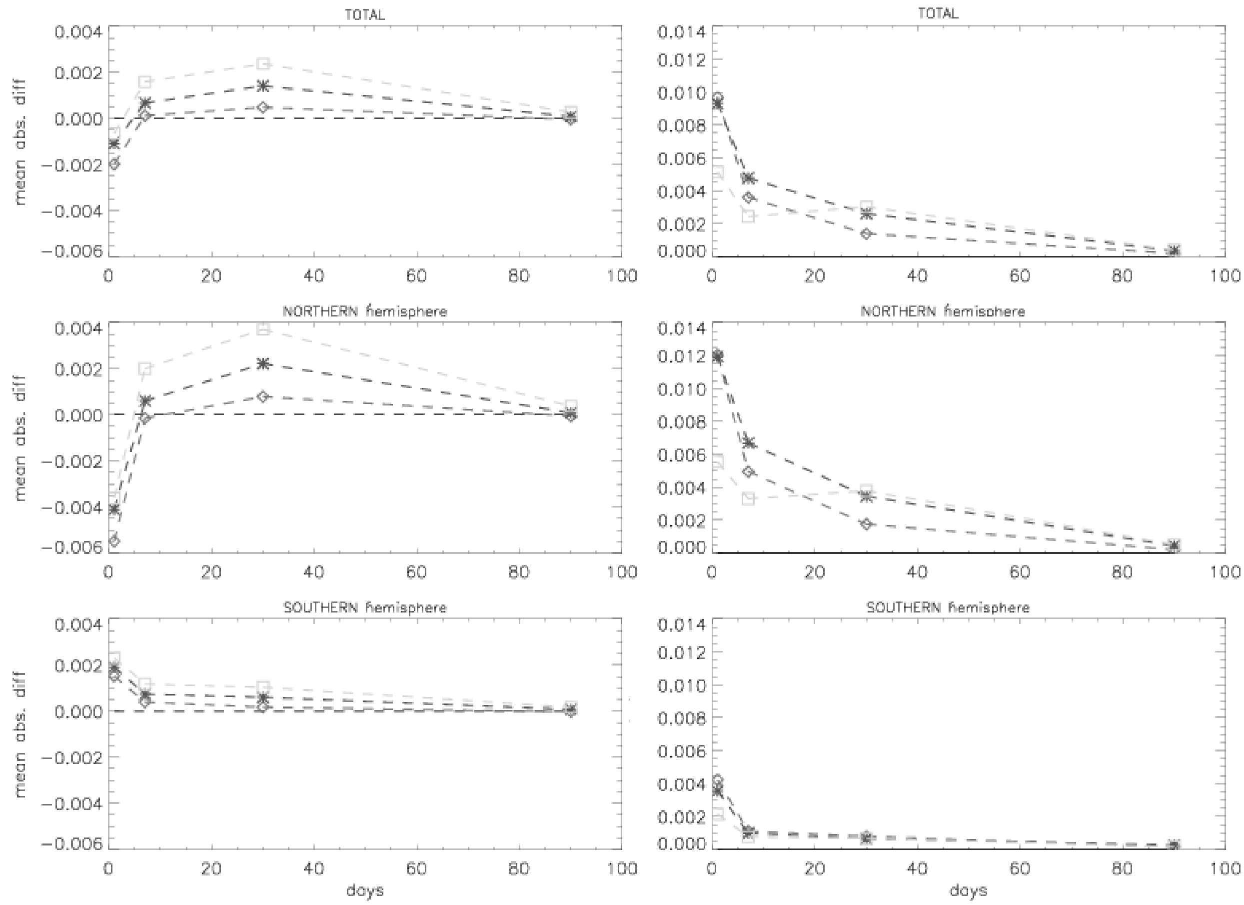

Figure 8. Mean absolute error and standard deviation as a function of time averaging for G1, G2 and G3

In an attempt to generally examine the distribution and characterization of the errors, the absolute difference of the seasonal AOD sampling bias at each G1 grid point is represented on figure 9(a). By measuring and contrasting over and underestimations of the satellite AOD sampling results with the true model values of AOD we intent to locate patterns of behavior and confirm that there are not any biases on the calculating procedure.

Figure 9(b) shows the corresponding true (model) values of AOD. We choose to show the winter results, being the conclusions drawn from the rest of the seasons equivalent. Winter is assumed to be three months starting the 1st of January until the 31st of March.

Correctly, the distribution of over and underestimation of the satellite sampling seems not to respond to seasonality or temporality. Likewise, there seems to be an equivalent number of over and underestimated grid cells. A certain pattern is observed following the satellite orbit footprints, such as to increase the sampling bias either positively or negatively at grid cells with overlapping footprints. This is reasonable; the higher then number of profiles sampled at a grid cell, the higher the error in comparison with the model.

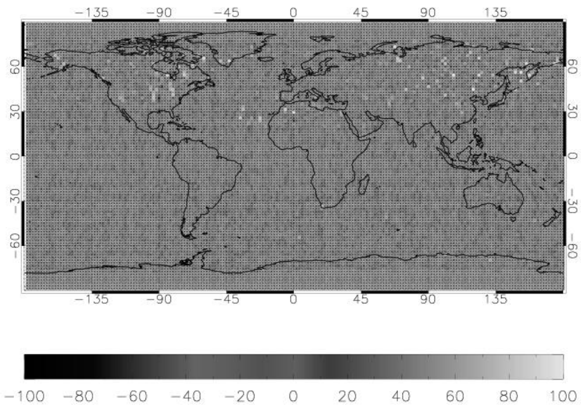

Figure 9(a). Winter 2001 sampling absolute diff. (%)

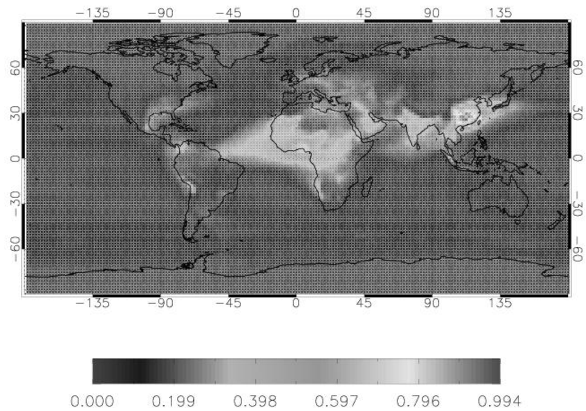

Figure 9(b). Winter 2001 averaged MATCH AOD

The highest discrepancies between the model and the satellite correspond to the steepest gradients of the model AOD values. Figure 10 shows the absolute AOD difference between the model's and CALIPSO values next to the model AOD data over the Eastern coast of China, which corresponds to the maximum gradient of AOD for this particular time period.

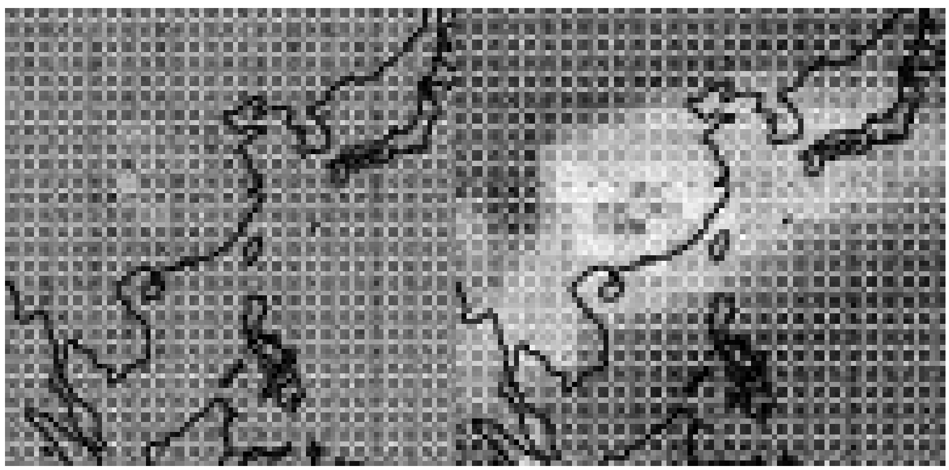

Figure 10. Detail of the AOD sampling study for the 2001 winter

6. CONCLUSIONS

CALIPSO nadir viewing lidar will provide us with a sparse field of measured aerosol properties. The uncertainties in using this subset to represent the full global field need to be ascertained. Using the MATCH climatology, specifically the aerosol optical thickness, we have sampled the satellite orbit and by means of zonal mean averages determined the required time and space averaging to reduce the error within our tolerance range. Regarding the results depicted in this article, weekly averages represent over a 60% error reduction in comparison with daily means but still hold a 10% global variance. Monthly zonal averages of aerosol optical depth reduce the error between true values and satellite profiles in comparison with daily means by 84% for G1 and G2, assuring consistent results throughout the planet. This error margin represents a maximum 17% discrepancy from the true AOD. The same monthly averaging results in a two-thirds reduction of the sampling uncertainties for the G3 grid.

Southern hemisphere has fewer aerosol sources than the Northern hemisphere, propitiating wider zonal averaging. Being the error variance negligible, there is no significant error reduction for time units longer than a week. Unless a uniform time and zonal averaging is required, these results need to be considered and a wider time and zonal averaging can be applied in the Southern hemisphere areas of the planet.

Further work needs to be done in the uncertainty study of regional averaging, as well as further research for individual aerosol species and other optical parameters such as the extinction coefficients.

7. ACKNOWLEDGEMENTS

Communication and partnership between agencies, the openness of the research world is undoubtedly the engine that propels our advancement in science. The authors of this paper have a debt of gratitude with D. Fillmore from NCAR, for diligently facilitating access to MATCH's climatology results and for agreeing on its use and publication.

As the launch date approaches for CALIPSO, work seems to multiply and our team generously devotes time and effort, fueled by anticipation and optimisms. We would like to thank Kathy Powell for her guidance and always finding time to make anything available to us.

8. REFERENCES

1. David M. Winker, Jacques R. Pelon, M. Patrick McCormick. "The CALIPSO mission: spaceborne lidar for observations of aerosols and clouds." Proc. SPIE **Vol.** 4893, p.1-11, Lidar remote sensing for Industry and Environment Monitoring III. Upendra N. Singh, Toshikasu Itabe, Zhishen Liu; Eds. 2003.

2. M. Santa Maria. "A radiative model for the study of the feedback mechanism between photolytic aerosols and solar radiation". Virginia Tech Publication Department. Master Thesis Dissertation, 2001.

3. P.J. Rasch, M.Lawrence. "Recent development in transport methods at NCAR". 1998.

4. Fillmore, DW , WC Collins and AJ Conley. "Aerosol direct radiative forcing – estimates from a global aerosol climatology constrained by MODIS assimilation", Journal of Geophysics Research, to be published, 2004.

5. Isaaks, E. H., and R. M. Srivastava. *An Introduction to Applied Geostatistics.* pp. 59, Oxford University Press, 1989.