Atmos. Chem. Phys. Discuss., 9, S2306–S2312, 2009
www.atmos-chem-phys-discuss.net/9/S2306/2009/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Data assimilation of CALIPSO aerosol observations" *by* T. T. Sekiyama et al.

T. T. Sekiyama et al.

Received and published: 27 May 2009

Dear anonymous referee#1:

Thank you very much for your fruitful comments on our manuscript. The responses to these comments are as follows. We hope these are satisfactory ones for you.

> 1. GEMS does not only focus on the assimilation and forecast of sea-salt, dust, organic and black carbon aerosols. So, on P5787L2, it could be included ", among others," between "forecast" and "sea-salt".

When the manuscript is revised, we change this sentence following your comment.

> 2. P5787L26: ". . . probably contain retrieval error". Can we have a bit more precision about this retrieval error?

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

In the fields of remote sensing, the retrieval error generally arises from 1) the difference between the actual vertical distributions of atmospheric components (such as temperature, pressure, humidity, concentrations) and their standard profiles for the calculation of radiances and 2) mathematical iteration processes to acquire a maximum likelihood estimate, when targeting variables (such as temperature and particle concentration) are retrieved from "raw" observational data. The retrieval error is one of the severe problems for data assimilation because the data quality deteriorates before the data are assimilated. In order to avoid the retrieval error, for example, the most advanced weather forecast centers in the world have recently been using "raw" radiance data measured by satellites, not retrieved temperature, for upper-air temperature assimilation.

> 3. P5788L7: change "532/1064" by "532 and 1064".

When the manuscript is revised, we change this sentence following your comment.

> 4. P5788L8-9: ". . . and these values are not contaminated by retrieval errors due to low-accuracy retrieval algorithms". This is no very clear to me. Do you mean that the retrieval algorithm does not generate errors nor amplify the instrumental error? Please, clarify.

We mean that the total attenuated backscattering coefficients do not include retrieval errors because the coefficients were directly measured and have not been processed by the retrieval algorithm. (Of course, instrumental errors are included in the measurements.) We have to apologize for the unclarity of this sentence. When the manuscript is revised, we change this sentence into "These values are not contaminated by retrieval errors because they were directly measured and have not been processed by low-accuracy retrieval algorithms."

> 5. P5788L14-28: This needs some more details or citations. > (a) OI and 3D-Var methods do not rely on the assumption mentioned in the manuscript but these assumptions are made in practice. I would write ". . . 3D-Var, practically assume . . . "

ACPD

9, S2306–S2312, 2009

When the manuscript is revised, we change this sentence following your comment.

> (b) A reference at the end of ". . . and temporally stationary" would be welcome.

When the manuscript is revised, we change this sentence following your comment.

> (c) What do you mean by ". . . 4D-Var implicitly evolves . . . "

Flow-dependent data assimilation, such as 4D-Var and EnKF, needs background error statistics. However, 4D-Var does not calculate the background covariance matrix directly. Instead, 4D-Var estimates matrixes that mathematically contain the background covariance information therein. You need enormous quantity of matrix calculation if you take the background covariance matrix anew from only 4D-Var results. In contrast, EnKF always calculates the background covariance matrix. We described this situation as "EnKF explicitly provides ..., 4D-Var implicitly evolves ...".

> (d) 4D-Var assumes model linearity during the assimilation window and not during "its iteration procedure".

When the manuscript is revised, we change this sentence into "Additionally, the 4D-Var involves the complexity of constructing linearized operators or adjoint matrixes, which are unnecessary for the EnKF" following your comment.

> 6. P5789L20: replace "of 1/3-1 km" by "between 300 m and 1 km"

When the manuscript is revised, we revise this sentence following your comment.

> 7. P5790L3: Can you tell us what do you mean by a "direct assimilation"

Previously on aerosol studies, retrieval data (such as extinction coefficient and aerosol concentration) have often been assimilated to avoid complexity in preparation of observational operators, although the data quality deteriorates through retrieval processes. In contrast, "raw" data were assimilated in our study, instead of retrieval data. The "raw" data are directly measured by CALIPSO, and have not been processed secondarily. Our data assimilation excludes retrieval data (= indirect data), and therefore, we

ACPD

9, S2306–S2312, 2009

Interactive Comment

call it "direct assimilation".

> 8. P5790L22: It is mentioned that "selected data were horizontally and vertically averaged to the model resolution". Do you mean to the model grid? Moreover, I could understand that high resolved profiles be averaged to the model vertical levels (while I would use the term "smoothed") but not horizontally. Do you mean that using several orbits, data are mapped on the model grid points? Please add some clarifications or, better, show us a plot with the original CALIPSO data and the transformed assimilated data. Finally, I suggest to include a panel in Fig 1 showing the assimilated data, i.e., the data averaged on the model grid with CAD score less than -33.

Vertically, we averaged high resolved profiles of measurements to the model levels. The vertical model levels are defined in a hybrid sigma-pressure coordinate which is depend on meteorological conditions especially in the troposphere. Therefore, CALIPSO sigma coordinate is not smoothly connected with the model coordinate, and the correspondence of CALIPSO coordinate to model coordinate varies moment to moment. Horizontally, the measurements with very high resolution along an individual orbit are chopped up "model's latitudinal resolution" in length, and then the chopped measurements are averaged for each length. The geographical coordinate (i.e., x, y, z = longitude, latitude, altitude) of each averaged data point is the "gravity center" of all measurement points included in each 2-dimensional (latitude vs. altitude) area. [CALIPSO is a lidar instrument, so the longitudinal width of each measurement is virtually zero.] Generally, each 2-dimensional area is not filled up with measurements with CAD scores less than -33. Consequently, the geographical coordinates of averaged data points do not agree with those of model grid points at all. This is the reason that we did not say, "be smoothed" or "be mapped on the model grid." The CALIPSO data were merely averaged to the size of approximately model resolution in order to decrease the data density. However, it is not essentially important for 4D-LETKF whether the data coordinates exactly agree with model coordinates. Instead, the density (or spatial representativeness) of assimilated data is useful information for data assimilation developers.

ACPD

9, S2306–S2312, 2009

Thus, we described the number of data points per day at P5590L25. In order to make some clarifications, we change the sentences P5590L20-L26 to "These measurements were selected only when the CAD score was less than or equal to -33, and then the selected measurements were horizontally and vertically averaged along each satellite orbit to approximately model resolution prior to data assimilation. This data selection markedly decreased the number of measurements used for data assimilation. After selection and averaging, the total number of B532, B1064, and d532 measurements to be assimilated was 15,000-25,000 points per day in the global troposphere. The geographical coordinate (i.e., longitude, latitude and altitude) of each data point is the barycenter of averaged measurements."

> 9. P5791L17: Can you shortly tell us what a "serial assimilation" is.

Data assimilation methods are classified into two categories. One is variational methods (such as 3D-Var and 4D-Var), the other is sequential methods. Variational data assimilation can deal with all observational data at once. Meanwhile, sequential data assimilation must deal with observational data one by one, basically. Kalman Filter is mathematically categorized as a sequential method. Therefore, observation data are serially assimilated with primitive EnKF techniques. However, it is inconvenient. Then many people have improved EnKF to deal with all observation data at once.

> 10. P5793L15-16: What is the "multiplication spread inflation parameter"?

Hunt et al. (2007) said, "In practice, an ensemble Kalman filter that adheres strictly to the Kalman filter equations may fail to synchronize with the "true" system trajectory that produces the observations. One reason for this is model error, but even with a perfect model, the filter tends to underestimate the uncertainty in its state estimate. Regardless of the cause, underestimating the uncertainty leads to overconfidence in the background state estimate, and, hence, the analysis underweights the observations. If the discrepancy becomes too large over time, the observations are essentially ignored by the analysis, and the dynamics of the data assimilation system become

ACPD

9, S2306–S2312, 2009

decoupled from the truth.

Generally this tendency is countered by an ad hoc procedure (with at least one tunable parameter) that inflates either the background covariance or the analysis covariance during each data assimilation cycle. ...(snip)... "Multiplicative inflation" instead multiplies the background covariance matrix (or equivalently, the perturbations of the background ensemble members from their mean) by a constant factor larger than one. "Additive inflation" adds a small multiple of the identity matrix to either the background covariance or the analysis covariance during each cycle."

The multiplication spread inflation parameter in the manuscript is "a constant factor larger than one" mentioned here.

> 11. P5793L16-17: By "The assimilated model variables/parameters", do you mean the system control variable, i.e., the variables optimized by the assimilation system?

We mean that "variables" are prognostic variables in simulation models and "parameters" are poorly known parameters in simulation models. Classically, data assimilation was able to deal with only prognostic variables. However, state-of-the-art data assimilation methods, 4D-Var and EnKF, can analyze not only prognostic variables but also poorly known parameters. In this case, the EnKF deals with all of the poorly known parameters as long as there is a correlation between the parameters and the prognostic variables.

> 12. P5793L18: ". . . and dust emission factors". Does is mean that, in addition to optimize the model state, the emissions are also optimized? In such a case, this would mean that your system is combining data assimilation and inversion of emission. If true, this would be mentionned earlier in the abstract and in the introduction.

Yes, the emissions are optimized. However, this is not a classic way of inversion. Our data assimilation is not categorized as a so-called "inversion method." State-of-the-art data assimilation systems, 4D-Var and EnKF, can optimize not only prognostic

ACPD

9, S2306–S2312, 2009

variables (=model state) but also poorly known parameters (=emission factor, etc). Emission optimization is just a part of data assimilation for these systems.

> 13. Sec 3.2: Several times, it is mentionned that the "4D-LETKF assimilation results" perform better that the reference model run. While this is true, this intents to say that this occurs thanks to the 4D-LETKF methods. Since you do not compare assimilation results of 4D-LETKF with another method, I would suggest to write only "assimilation results" (i.e., dropping "4D-LETKF"). Moreover, the good results are at least as good as they are thanks to the good quality of the CALIPSO data as due to the choice of the assimilation method.

When the manuscript is revised, we drop "4D-LETKF" from "4D-LETKF assimilation results" following your comment. We deeply appreciate the CALIPSO team providing their data. Without the CALIPSO data, we couldn't have made on this study. On the other hand, without 4-dimensional ensemble Kalman filter, value-added (= gridded, interpolated, extrapolated and aerosol-type-partitioned) data cannot be derived effectively from lidar "raw" data. This is a mutual benefit.

> 14. P5799L13: ". . . without retrieval errors". Same comment as above.

Please refer to the reply to #2.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 5785, 2009.