

NOAA/NASA/DoD Workshop on Satellite Data Assimilation

Ronald M. Errico
National Center for Atmospheric Research

George Ohring
National Environmental Satellite Data Information Service, NOAA

John Derber
National Centers for Environmental Prediction, NOAA

Joanna Joiner
Data Assimilation Office, NASA

ABSTRACT

A workshop on the assimilation of satellite sounding information using global forecast and climate models was held at College Park, MD, 23-25 August 1999. Topics discussed included: comparisons of assimilations of satellite retrievals versus satellite-observed radiances, planning for the use of advanced infrared sounders, the use of satellite sounding data affected by land surfaces, radiative transfer issues, and error characteristics of models and observations. The workshop concluded with a number of general and specific recommendations to advance the state of the art of assimilation of satellite sounding data.

1. Introduction

A workshop on the assimilation of satellite observations for use in global forecast and climate models was convened at the University of Maryland Conference Center in College Park, MD during 23--25 August 1999. It was hosted and supported by the United States National Oceanic and Atmospheric Administration with additional support from the National Aeronautics and Space Administration, and the Department of Defense. Organizers included Ronald Errico (NCAR), George Ohring (NESDIS), Joanna Joiner (NASA), Steven Lord (EMC/NCEP), and Marie Colton (ONR). The three sponsoring agencies had organized a similar workshop in April 1998, which covered assimilation of a wide variety of satellite data products (Errico, 1999). The present workshop focussed

on the assimilation of information from atmospheric sounding instruments. Approximately 50 remote sensing scientists and data assimilation experts from the U. S. and abroad participated. Topics discussed included: comparisons of assimilations of satellite retrievals versus satellite-observed radiances, planning for the use of advanced infrared sounders, the use of satellite sounding data affected by land surfaces, radiative transfer issues, and error characteristics of models and observations. The format included 3--4 short invited talks on each topic with equal time available for questions and discussion.

2. Comparisons of Retrieval and Radiance Assimilations

The purpose of this session was to clarify the differences between assimilating satellite sounding information indirectly using retrievals of temperatures and humidity and directly using radiances. Presentations were made by Joanna Joiner, Clement Chouinard (RPN), Richard Renshaw (UKMO), and Brett Harris (BMRC).

One frequent difference between assimilating retrievals and assimilating radiances is due to the different background (first guess) fields each typically uses. The latter approach almost invariably uses short term forecasts from a numerical weather prediction/analysis system while the former sometimes use generally less accurate background information from an air mass classification system or climatology. Such differences can profoundly impact results because the information content of satellite observations is rather limited due to the vertical breadth and redundancy of spectral information and the limited accuracy of both instruments and radiative transfer models.

If retrievals are produced using backgrounds that are identical to or correlated with the background information used by the subsequent data assimilation system, then when the retrieval is treated as an observation to be assimilated, its errors are in general correlated with those of the background. This violates the useful and common assumption that observation and background errors are uncorrelated and requires consideration of additional, cross-correlation terms within an assimilation's variational framework. These can render the assimilation poorly posed or intractable. In a 1-dimensional (vertical) framework, neglecting these terms actually has small impact on most vertical structures, but in a 3-dimensional framework their neglect may cause greater degradation of the optimality of the assimilation result.

The direct assimilation of radiances permits careful monitoring of some observational and radiative transfer model error statistics. For example, routine monitoring of innovation statistics, determined from accumulated differences between observed radiances and corresponding simulated radiances produced by applying a radiative transfer model to background fields interpolated to the observation locations, reveals systematic errors (biases) that are scan-angle dependent. These biases are large and must be first identified and subsequently removed if the remaining portion of the satellite data is to have a positive impact on the assimilation system. Some of these biases depend on not only viewing angle but also viewing direction, and may in fact indicate effects of

electromagnetic interference created due to the instrument configuration on the observing platform. These biases would likely be less easy to detect in retrievals produced using less-accurate background fields. The biases in retrievals should also be examined and corrected if detected

Most of the issues regarding optimal assimilation of either retrievals or radiances are common to both techniques. For this reason, some of the debate regarding the superiority of one technique over another is a distraction: What is best for one operational setting, instrument, or application may not be best for another. The trend at operational, numerical weather prediction centers is towards direct assimilation of radiances, based on the successes at both NCEP and ECMWF. The trend is also towards using less-processed data (i.e., no limb adjustment or cloud clearing). The careful consideration of statistics, selection of background, and treatment of biases have been revealed to be critical for the optimality of either technique.

3. Use of Advanced Sounders

The two advanced sounders that will be providing data soon are the Atmospheric Infrared Sounder (AIRS) on NASA's EOS PM spacecraft, scheduled for launch in 2000, and the Infrared Atmospheric Sounding Interferometer (IASI) on EUMETSAT's METOP-1, planned for launch in 2003. These instruments provide radiance estimates for thousands of channels and require new considerations for handling their data volumes. Presentations on possible handling of these new data were offered by Florence Rabier (Meteo-France), Mitch Goldberg (NESDIS), and Joel Susskind (NASA).

For those operational and research centers with adequate computer power, it may be possible to assimilate most pertinent channels for intended applications, but less-equipped centers will likely need to do more extensive data thinning, either in terms of channels or spatial density. This assumes that most of the data from the advanced sounders will actually be available in real time. Ways to realize this availability are currently being explored.

An initial thinning approach being pursued at NESDIS is to screen out (obviously) cloud-contaminated radiances to produce clear radiance data sets for dissemination to users. Cloud contaminated infrared radiances are already ignored by most current assimilation systems, and therefore this thinning would not affect those systems until ways of extracting more useful information from such radiances are developed. Other approaches include creation of super-channels from set of spectral channels with similar weighting functions or generation of principal components of some description of the information content within the observations.

Presently, with the horizontally broad functions used to characterize background error correlations, spatial thinning of rather dense satellite observations does not exclude much additional information that assimilation systems could otherwise absorb. Radical

changes to those background error statistics may be required before additional benefit of the already available dense observations can be realized. Typically, the deep vertical scales that satellites effectively measure are associated with structures having large horizontal scales also, and therefore the effective horizontal smoothing of the observations by the assimilation system is appropriate in an average sense. It may be, however, that when flow-dependent statistics are incorporated into assimilation systems, rather than time-independent statistics as currently employed, the capability of usefully ingesting denser observations in some synoptic situations may be realizable.

Each observing system requires much detailed consideration and software development before and after it can be incorporated into an assimilation system. Each system likely requires its own peculiar tuning, quality control, and monitoring, but much experience and even software can be shared. Examples of such cooperation within Europe are the assignment of EUMETSAT staff to ECMWF and the inclusion of data assimilation experts in the IASI project at Meteo-France. ECMWF and NCEP have shared personnel developing TOVS radiance assimilation, and more recently, NASA's Data Assimilation Office (DAO), NESDIS, and NCEP have committed themselves to cooperative efforts regarding the assimilation of satellite observations. Radiative transfer code has been freely exchanged between many organizations, including NRL, DAO, NCEP, ECMWF, UKMO, NESDIS, and other researchers. There are fewer examples of cooperation between operational numerical weather prediction centers and either data producers, instrument developers, or university researchers in satellite sounding utilization. Cooperation among these other groups must be encouraged if progress is to be timely and resources are to be expended wisely.

4. Use of Satellite Sounding Data Affected by the Surface

Satellite soundings affected by land surfaces are currently utilized little, for the good reason that the surface contributions may be grossly uncertain. This does not mean that there is only poor quality information over land, but instead that more effort is required to use it properly. Also it means that there is an opportunity to estimate surface properties. These issues were presented by Fuzhong Weng (NESDIS), Thomas Kleespies (NESDIS), and Albin Gasiewski (ETL).

Surface radiance depends on both the surface emissivity and skin temperature. What a satellite observes of the surface is further affected by atmospheric transmission of the radiation. For spectral channels whose weighting functions peak high in the atmosphere, surface contributions are negligible, but for the lower-peaking spectral channels, surface contributions can be significant.

For the ocean surface, good models have been developed for relating infrared and microwave emissivities to surface wind speed and satellite viewing angle. With the good quality of available information on ocean surface temperatures, both infrared and microwave sounding channels can therefore be assimilated over the oceans. It is possible

to use such observations to estimate surface wind speed and, given multiple viewing angles, wind direction also.

Assimilation over land surfaces poses greater problems due to its larger variability of both emissivity and skin temperature. Infrared emissivities of land surfaces vary from about 0.9 to 1.0, and microwave emissivities vary from about 0.4 to 0.98. These variations exist not only due to differing land types, but in the case of microwaves, also due to surface water content or the phase of cropland utilization: a field of corn can have a very different microwave emissivity than a field of dried stubble. Surface skin temperature can be strongly affected by viewing angle as the portion of surface shaded by trees or clouds changes.

In the short term, perhaps the best way to tackle the problem of retrievals over land surfaces is to focus on what types of surfaces may be easier to consider and may have a significant impact on the results. For example, over some deserts, emissivity may be more spatially and temporally uniform than over croplands. Ice, which can pose problems at near freezing temperatures, may also be more amenable to consideration under some other conditions. A partial solution to this problem is better than none, and the surface types just mentioned cover a significant portion of the earth that is otherwise only poorly observed.

It is also possible to develop models of surface emissivity as functions of land surface type and surface wetness. Dynamic mapping of emissivity could make use of satellite measured vegetation indices. With surface emissivity known, skin temperature can be estimated from atmospheric window measurements, as is done for sea surface temperature. For use in assimilation systems, these estimates must be sufficiently accurate, regarding both the biases and random components of their errors.

5. Radiative Transfer Issues

Radiative transfer models are key components in the optimal assimilation of satellite soundings. Computationally fast models are required for real-time use. These are designed by parameterizing line-by-line models for application to the channel suite for each type of instrument. Issues regarding their development and accuracy were discussed by Larry McMillin (NESDIS), L. Larrabee Strow (U. MD), Louis Garand (AES), and Norm Grody (NESDIS).

Several steps are required to develop a fast radiative transfer code. Basic spectroscopy must be performed in a laboratory, to determine how different chemical species absorb radiation, including details of absorption-line shapes. Models describing the radiative transfer for each spectral line must be developed. Then the convolution of an instrument's channel response and the spectral absorption/transmittance must be determined. Finally, this result must be parameterized so that the computations can be performed at a reasonable rate for processing data in real time.

All these steps are sources of error as well as information. The root-mean squared errors in each step may be several percent of the signal so that the net error may be as large as 10% of the signal. The variance of this error combines with that of the instrument to yield the inverse of the optimal weight given to the observation in the specification of a cost function to be minimized in a data assimilation procedure.

For TOVS and AIRS, the fast radiative transfer models perform well. Significant spectroscopy for most important frequencies has already been preformed (weak water lines in the infrared may be an exception). Fast monochromatic codes have also been developed that allow for quicker creation of line-by-line databases.

Currently concentrations of gases such as CO₂, CH₄, and N₂O are treated as constants in some radiative transfer codes, although it is known that they have significant spatial and temporal variations. Other simplifications that contribute to errors are the Lambertian assumption for the surface, errors in Jacobians, neglect of reflected down-welling radiation, and neglect of scattering. Most of these issues are now being addressed by the community.

The community that performs laboratory spectroscopy and that develops and maintains line-by-line and fast radiative transfer models is shrinking. Some of the work is tedious and unexciting. Few are trained in such work now and it is difficult to fill positions when they are available. This should be of great concern to all who rely on good, global atmospheric analyses.

6. Error Characteristics of Models and Data

The data assimilation problem is fundamentally a statistical one because all the information provided by models, observations, and prior analyses have unknown errors, that can only be described statistically. Characteristics of these errors were described by Kim Richardson (NRL), Tony McNally (ECMWF), Erik Andersson (ECMWF), Robert Aune (NESDIS), and John Derber (EMC).

The most accurate source of information about the current state of the atmosphere, prior to considering any current observations, is a short-term numerical forecast begun from an atmospheric analysis determined by previously considered observations. Essentially, a forecast model is used to extrapolate previous information to the present time using dynamical and physical relationships. Information extrapolated in this way is as accurate, in a root-mean-squared sense, as that obtained from most observation systems alone.

The primary difficulty with specifying the background errors is that they are spatially correlated by the dynamical and physical processes acting in both the atmosphere and model. Currently, our best estimates of these correlations are rather broad and time independent. They imply that background errors are dominantly on scales of hundreds of

kilometers, and these are the scales that the assimilation therefore tries to correct. Consequently, data at much higher horizontal resolutions are effectively smoothed. This will occur as long as our estimated error correlations are broad, which will likely remain the case on average, although for some synoptic situations they may be narrower. Not until we have flow-dependent error statistics, rather than time-independent ones, will it be possible to actually retain small-scale information from high-resolution data under some conditions.

Currently, there are several efforts to generate flow-dependent error statistics. These include several kinds of ensemble, sub-optimal, and parameterized Kalman filter techniques. The problem is complicated, however, and fraught with computational difficulties. Much fundamental theoretical work is likely also required.

The existence of both model and observation bias is another problem. Gross biases are typically removed during model development and instrument calibration. Remaining, more subtle but still significant biases may be revealed by computing statistics of differences between observed radiances and corresponding simulated radiances. The latter are produced at each observation location by applying a radiative transfer model to estimates of the atmospheric profiles that are independent of the observations being considered, typically as provided by a short-term forecast in lieu of more accurate estimates. Although only differences between the two types of biases are explicitly revealed by this method, in some instances it is possible to distinguish one from another by examining correlations of the biases and comparisons with nearby observations. In other instances, however, it is difficult to determine whether a bias is due to an error in the model background or observation. Yet, it is important that the error be corrected properly: an observation bias can simply be subtracted from an observation, but if the bias is actually in a model generated background, its subtraction from the observation would fail to remove the bias from the analysis.

Biases that depend on the scan angle of the satellite observing system can occur due to viewing angle dependent errors in the radiative transfer model, the observing instrument, or the data processing. Such biases are less likely a result of errors in the forecast used to produce the estimated atmospheric profile. Much of the development in assimilating satellite data is consumed in estimating their biases.

There is now much exchange of information among users of TOVS data. Communication between those performing data assimilations (as end-users) and those involved with ground calibration, however, should be enhanced. In particular, estimates of bias dependent on scan angle suggest some basic instrument calibration problems or errors in the radiative transfer model or parameters.

Observation System Simulation Experiments (OSSEs) are performed to evaluate the impact of future observing systems on weather predictions. OSSEs should be well designed and conducted. They should be performed with consultation of experts at the operational centers to ensure that realistic, current estimates of error statistics are used. How realistic the models or statistics must be depends on the types of questions being

asked. Answers to qualitative questions, such as "How do errors in initial specification of internal unbalanced structures affect forecasts?" may require less realistic simulations, but answers to quantitative questions, such as "How much does this new data type improve current forecast accuracy?" require very realistic simulations. The answer to the latter question depends critically on the other data types considered with it and can change as the model, assimilation, or observing systems change. With a good OSSE, however, the explanation, and thereby the generality of the results can also be explored. Performing OSSEs using operational data assimilation systems requires the preparation of many of the components that are necessary for eventual operational use, thereby significantly shortening the time from satellite launch to data utilization.

The background fields and radiative transfer models used by a data assimilation system also provide estimates of radiances that are otherwise only imperfectly observed. Such estimates can be optimally combined with observations, by properly accounting for the error statistics of all the pieces of information. Mathematically, the problem of producing an optimal analysis of radiance at observation locations is identical to that of assimilating data to produce estimates of gridded fields, and therefore not much additional work should be required to obtain such estimates from an existing data assimilation system. Such a-posteriori estimates can enhance sorely needed diagnostic examination and verification of the observing system.

7. Workshop Recommendations

Recommendations of participants at the workshop fall into two broad categories: general and specific:

Some general recommendations are:

- Background error statistics must be improved, most notably by rendering them flow-dependent.
- Model and observation biases must be better estimated and distinguished.
- Health of the science requires that ways be found to encourage training and work in critical areas such as radiative transfer and data assimilation.
- Satellite data providers and data assimilation groups should work together to realize the full potential of satellite data in models
- Some guidelines on performing and evaluating OSSEs should be established and publicized.
- The U.S. should apply the best super computers to the problem of optimal data assimilation.

- To be useful for numerical weather prediction, observational data must be available in real time; i.e., within 2-3 hours of being collected.
- Data assimilation systems should be augmented to produce a-posteriori estimates of observations for diagnostic evaluation of the observing system.

More specifically:

- The fast radiative transfer code OPTRAN should be maintained and its improvement continue.
- NESDIS should be enabled to provide AIRS data in real time for use at operational centers.
- Ground calibration providers should communicate with end-users, particularly at operational centers, who are carefully examining their products and who therefore require detailed knowledge of the data processing and can provide important feedback regarding error characteristics.
- Land surface types most amenable to the sounding retrieval problem should be identified and the work first focused on retrievals and assimilation over those surfaces.
- Spectroscopic measurements required to utilize new instruments must be performed, the necessary line by line models developed, and results from the latter incorporated into fast models. Although this work is critical, its current support in terms of facilities and experienced personnel appears very tenuous.

Acknowledgements

Although the writing of this report was primarily performed by its authors, all of the workshop participants contributed through their presentations, questions, discussion, and review. The authors thank them all and the supporting agencies.

REFERENCES

Errico, R.M., 1999: Workshop on satellite data assimilation. *Bull. Amer. Meteor. Soc.*, **80**, 463-471.