Current Status and Future of Satellite Data Assimilation

John C. Derber and Andrew D. Collard

Environmental Modeling CenterNCEP/NWS/NOAA
Camp Springs MD, USA
John.Derber@noaa.gov
Andrew.Collard@noaa.gov

Abstract

The use of satellite data has become a very important component of modern data assimilation systems. However, care must be exercised to properly use the information in the satellite data. The ability to accurately simulate observations from the analysis variables is essential. If the data is not well simulated, it cannot be reliably used within the system. The use of the data and the development of the forward models requires good knowledge of the instrument characteristics. The satellite data can have significant biases. These biases (if not from the background field) should be removed prior to the use of the data. The specification of the observation error covariance matrix also can play an important role in the use of the data. Finally, computational cost considerations provide a strong factor in the design of satellite assimilation systems. The future entails the inclusion of many new analysis variables, many new instruments and will require the enhancement of all components of the assimilation system.

1. Introduction

Over the last 20 years the importance of satellite data in operational data assimilation systems has continually increased. While some of the increase in the impact of satellite data comes from improved instruments, most of it comes from our increased ability to use the information in the observations. In this paper, we will discuss some of the characteristics of the satellite data, techniques for using the data and the possible future of satellite data usage.

The first question that arises is "what makes satellite data different from other observations?" There is not always a clear answer to this question, since many ground-based observations can have some of the same characteristics. Satellite data usually has a unique combination of:

- remote sensing of the observation (attempting to measure something at distance from the instrument)
- indirect observation of the quantities that you wish to measure (e.g. radiances rather than temperatures/moisture/etc.)
- one instrument making observations in many different locations
- a huge volume of observations
- data is often asynoptic
- instruments cannot be accessed for repair or re-calibration
- the instruments are very expensive.

Because of all of these aspects, just the design of the instruments can be a long, expensive proposition. Properly accounting for the details of the design can become quite important in our ability to extract useful information from the observations.

Most satellite instruments fall into 3 categories; active, passive and occultation instruments. The active instruments send a signal towards the earth and measures how the signal is returned to the instrument to infer information about the earth. A simple example would be a satellite based precipitation radar, which will send out a signal at a particular frequency and measure the return to estimate the properties of clouds and precipitation. One of the major limitations of active instruments is that sending the signal from the satellite requires energy which limits their lifetime.

Passive instruments rely on measuring radiation emitted from the earth's atmosphere and surface. The infrared and microwave radiance observations used extensively in operational data assimilation systems fall into this category. These instruments can fall into two different observing strategies, nadir and limb sounders. Nadir instruments tend to have good horizontal resolution while having poorer vertical resolution, while limb observing instruments generally have better vertical resolution and poorer horizontal resolution. These two types of instruments provide one of the examples of the importance of combining different types of observations to fully resolve the complete state.

Occultation instruments measure the impact of the atmosphere on a known signal being transmitted through the limb of the atmosphere. The transmitter can be the radiation from natural sources such as the sun, stars, or moon or from artificial sources such as the transmitters on the Global Positioning System (GPS) system. These observations can produce high quality measures of atmospheric state along the path from the transmitter to the receiver, as demonstrated by the example of the GPS Radio-occultation measurements (Healy, et al., 2007, Cucurull,2010).

2. Using satellite data

The assimilation of satellite data has been an area of great progress at operational NWP centres over the last 20 years. During this time many important lessons have been learned. First and most important is that one must treat satellite data very carefully. Without care in use of the observations, the volume of observations can easily result in a significant negative impact as easily as a positive impact. It is imperative to properly model the instrument characteristics for full use of the information to be possible. However, since all the instrument characteristics (e.g., spectral response functions) are not perfectly characterized and since our use of the data is limited by computational resources, it is safe to say that we are not using any component of the observing system (satellite or otherwise) to its full potential. So much of the improvement over the last 20 years has resulted from improving the details of the use of the data to improve the utilization of the data.

There has been much discussion about the use of raw observations vs. pre-processed observations in data assimilation (e.g., radiances vs. retrievals). Generally if one can account for the pre-processing in the quality control, background, and specification of observation errors, one can use either with equivalent results (within a linear system). However, accounting for the pre-processing in the quality control and in the specification of the observations errors can be difficult (distinguishing the signal from one bad channel in the retrieved profile), computationally expensive (communicating NxN matrices from the retrieval to the analysis) or made very difficult by nonlinearity in the transformation from the analysis to observation variables. For these reasons, it has generally been found to be better to use the observations with minimal processing.

The variational analysis problem can be stated as:

$$J = (x - x_b)^T B^{-1}(x - x_b) + (K(x) - y)^T (E + F)^{-1}(K(x) - y) + J_C \tag{1}$$

Where

x	= Analysis
x_b	= Background
B	= Background error covariance
K	= Forward model (nonlinear)
y	= Observations
E+F	= R = Instrument error + Representativeness error (including forward model error)
J_C	= Constraint term.

The second term in this equation for the use of satellite data is the most important term. The forward model K(x) transforms the analysis variable x into the same form as the observation y. The instrument and representativeness error determine the weighting of the difference between the simulated observation from the analysis variable and the observation. Each of these terms represents important components in the use of all observations, especially satellite observations.

The forward model (K(x)) which transforms the analysis variables into the form of the observations, can incorporate many different components. In 4d-var, it contains the integration of the forecast model from the analysis time to the observation time. When the analysis variables are different from the variables need to create the observations, (e.g., streamfunction and velocity potential), the forward operator can contain a transformation to the form necessary (e.g., u,v components of the wind). Finally, the forward operator contains the creation of the simulated observation. The creation of the simulated observation can be as simple as an interpolation to the location of the observation (when the analysis variable is the same as the observation), or as complex as an integration of a fast radiative transfer model to simulate the observed brightness temperatures or the simulation of the bending angle for a GPS-RO observation. Whatever is contained in the forward model, the errors in the simulation of the observations should be kept as small as possible and should not be larger than the signal in the observations.

As an example throughout the paper, the satellite radiance assimilation problem will be used, since this type of data represents a large percentage of the available satellite observations. If the observations (y) are satellite radiances (usually in the form of brightness temperatures) then we can create a simulated observation (y) by integrating a radiative transfer equation (K(x,z)). The input (x) to the forward operator (K) in this case are profiles of temperature, moisture, ozone and various surface parameters such as skin temperature, emissivity and surface type. K is also a function of z which represents the unknown inputs of the radiative transfer such as methane profiles, aerosol profiles, cloud and precipitation. In general K is not invertible with more unknowns than degrees of freedom in x (and z). Satellite retrievals have been created to transform the observations into more standard meteorological variables. These retrievals basically fall into 2 categories, statistical and physical retrievals. The statistical retrievals are empirical regression equations developed by regressing the solution against the input values to get the most statistically likely relationship between observations y and the input variables x. The physical retrievals are essentially 1-D variational versions of equation 1. Similarly, the analysis problem in equation 1 is also an underdetermined problem and can be thought of as a 3/4D generalization of the physical retrieval problem with the inclusion of many different observation types.

If the errors in either the formulation of K(x,z) or the signal from the unknowns z are too large, the data cannot be reliably used and the observations should be removed in the quality control. For example, if the radiative transfer calculations do not incorporate clouds in the radiative transfer, the use of observations with a cloud signal in them is not advised. When observations are used that are not properly simulated, there is also a significant likelihood that the errors of different observations of this type will have significant correlations. If this correlation is not properly accounted for (which is often very difficult), it can result in a significant problem in the analysis. Also note that when the formulation is improved or variables are moved from z to x, more data may be used in the analysis and less data rejected by the quality control.

The forward operator can be nonlinear (for example wind speed and direction when the analysis variable is streamfunction and velocity potential), but it should be differentiable. For satellite radiances, the use of moisture signal tends to be slightly nonlinear, while the inclusion of cloud and precipitation can add strong nonlinearities. If the forward operator is nonlinear, then a nonlinear minimization algorithm must be used. The nonlinear minimization algorithms tend to be more expensive, more complex and a unique minimum is not guaranteed. Also, the inclusion of the nonlinear forward model will often slow convergence.

The satellite radiance observation instruments measure the upwelling radiation at the satellite. The upwelling radiation at the satellite results from emission and absorption over deep layers of the atmosphere and the surface of the earth. These layers are deepest for the microwave instruments. For the older infrared instruments such as HIRS, the layers are not quite as deep, but still cover a large percentage of the atmosphere. The high spectral resolution instruments such as AIRS and IASI are narrower still, but the measurements still come from a quite deep layer of the observation. The higher vertical resolution claimed from these instruments come from the de-convolution of the information in the many overlapping deep layers. Note that deep layers of the atmosphere are usually associated with large horizontal scales with some notable exceptions such as hurricanes. Thus, it is unlikely that this type of observations will give us much information on small horizontal or vertical scales.

To integrate the radiative transfer model for all instruments and channels can be quite expensive. For this reason, fast radiative transfer models have been developed which can perform the integration much faster than line-by-line models. Two of the most extensively used fast radiative transfer models are RTTOV (http://research.metoffice.gov.uk/research/interproj/nwpsaf/rtm/) and CRTM (ftp://ftp.emc.ncep.noaa.gov/jcsda/CRTM/CRTM_User_Guide.pdf). If one is solving eqn. 1 variationally, the adjoint or Jacobian of the radiative transfer is also needed. The adjoint and/or the Jacobian are provided along with the simulation of the observations by RTTOV and CRTM. Much can be learned about the instrument's measurement capabilities by examining the Jacobians, which give the partial derivative of the observation with respect to the input quantities (i.e., temperature, moisture, ozone, etc.). In Fig. 1, the brightness temperature Jacobians with respect to temperature for the HIRS/4 on METOP-A are shown. Note the deep layers over which the brightness temperature is impacted by changes in the input temperature profile. Also note the non-zero values at the surface for many of the channels. This indicates that a significant part of the signal will come from the surface, rather than the atmosphere.

In Fig. 2, similar plots are shown for the IASI channels closest to the centre of the HIRS/4 channels shown in Fig. 1. As expected there are many similarities between these plots, but the IASI Jacobians are a bit narrower in the vertical.

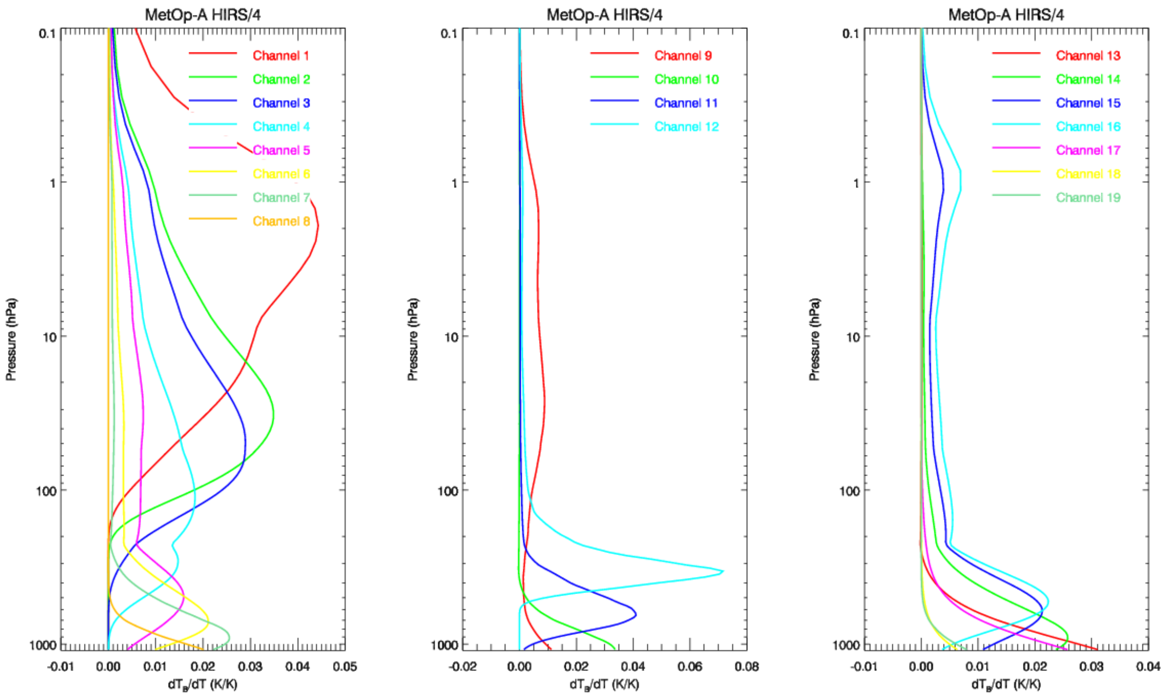

Figure 1: Temperature Jacobians for METOP-A HIRS channels for U.S. standard atmosphere.

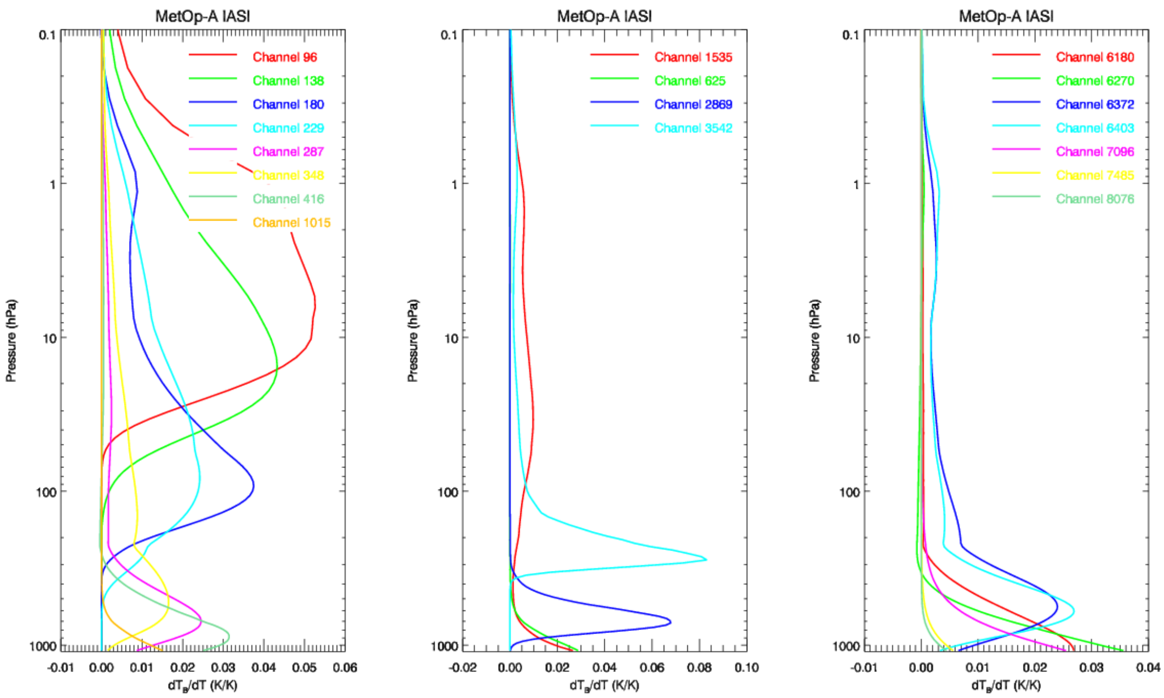

Figure 2: Temperature Jacobians for a selection of METOP IASI channels located near the centre of the METOP-A HIRS channels. These profiles were produced using a U.S. standard atmospheric profile.

In Fig. 3, the temperature Jacobians are shown for the AMSU-A on METOP-A. Note the much deeper Jacobians for the AMSU-A channels. Also note the much simpler structure of the weighting functions (almost Gaussian) for the microwave channels. The first 5 microwave channels and channel 15 are very strongly impacted by the surface signal.

Fig. 4 shows the Jacobians with respect to moisture for the same HIRS/4 channels as in Fig. 1. Note the change in sign in most locations. This means that as you increase the amount of moisture in a

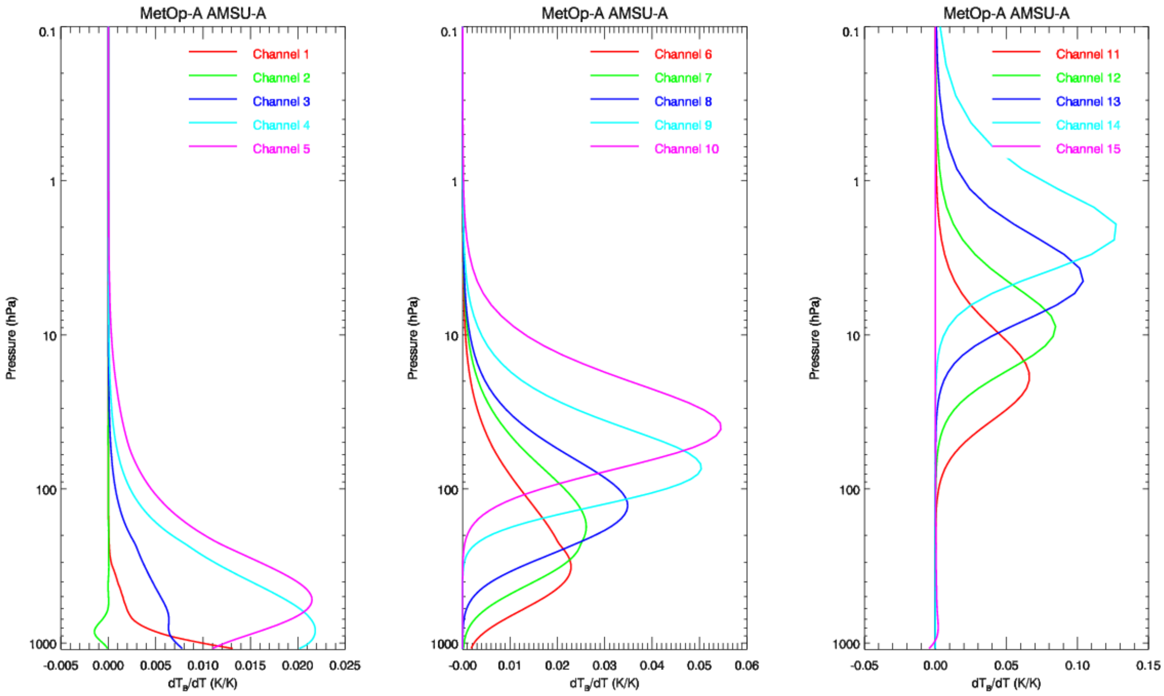

Figure 3: Temperature Jacobians for METOP-A AMSU-A channels for U.S. standard atmosphere.

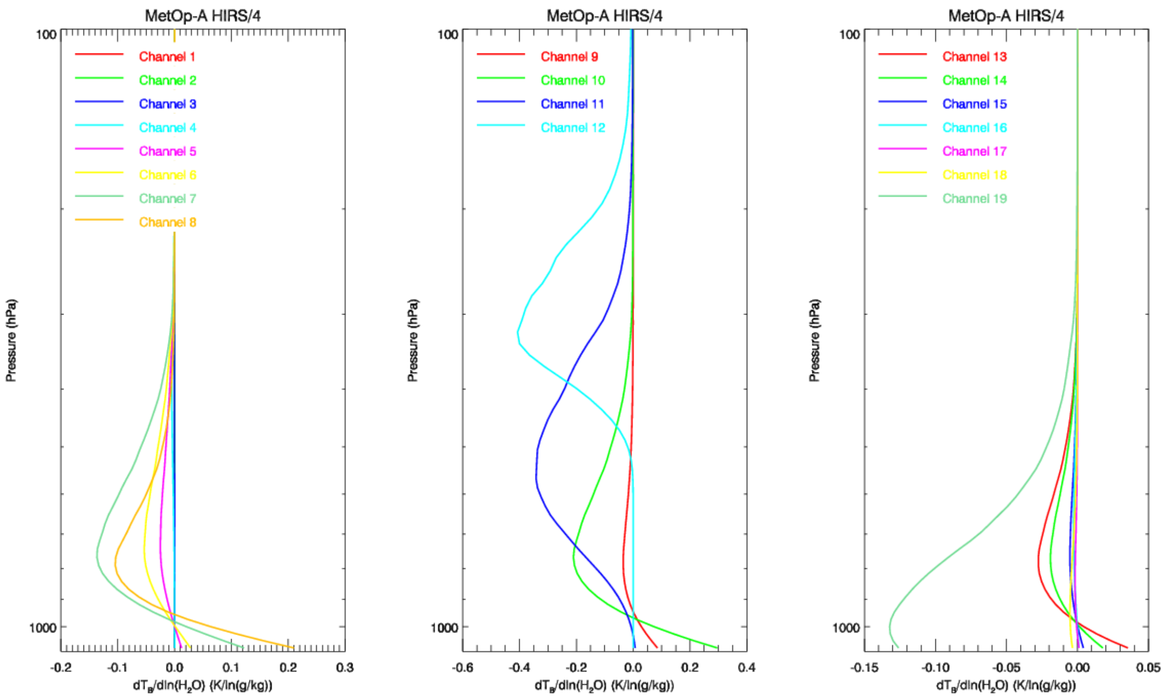

Figure 4: Moisture Jacobians for METOP-A HIRS channels

profile, the brightness temperature is likely to decrease. The decrease in temperature occurs because as the amount of moisture increases the atmosphere becomes more opaque and the instrument will be measuring higher in the atmosphere. Similar moisture Jacobians are found for the IASI channels shown in Fig. 2.

Fig. 5 shows the moisture Jacobians for the AMSU-A. Please note the change in scales between the panels. For channels 1-5 and 15 there are significant moisture sensitivities for AMSU-A. For channels

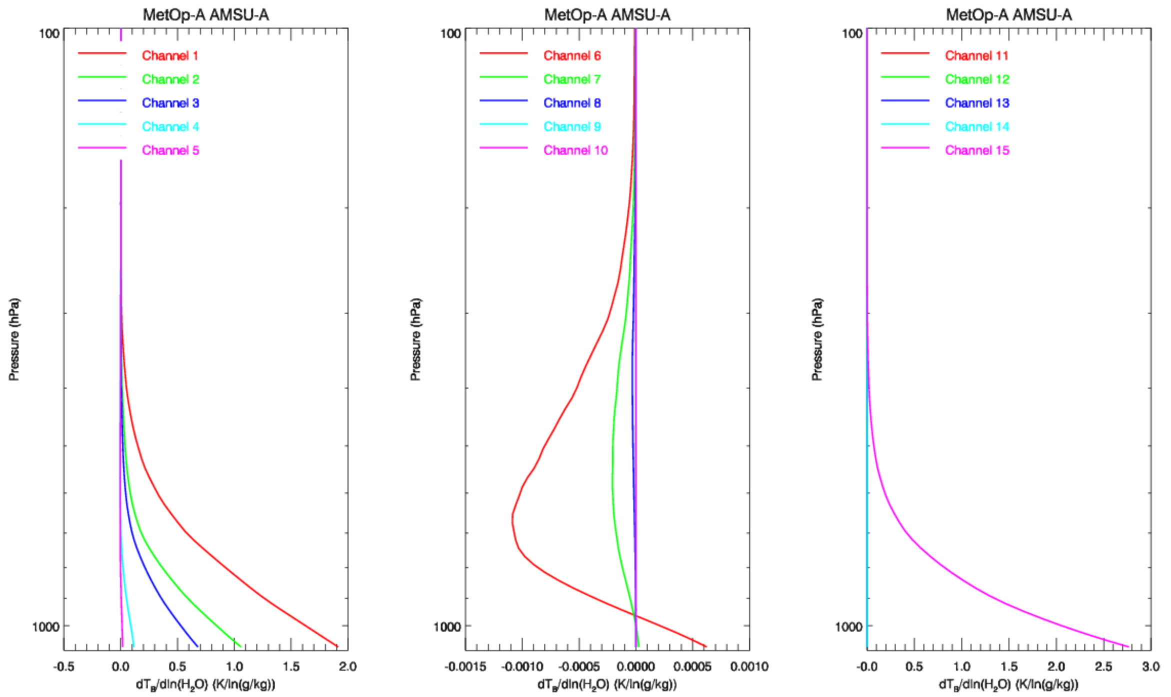

Figure 5: Moisture Jacobians for METOP-A AMSU-A channels

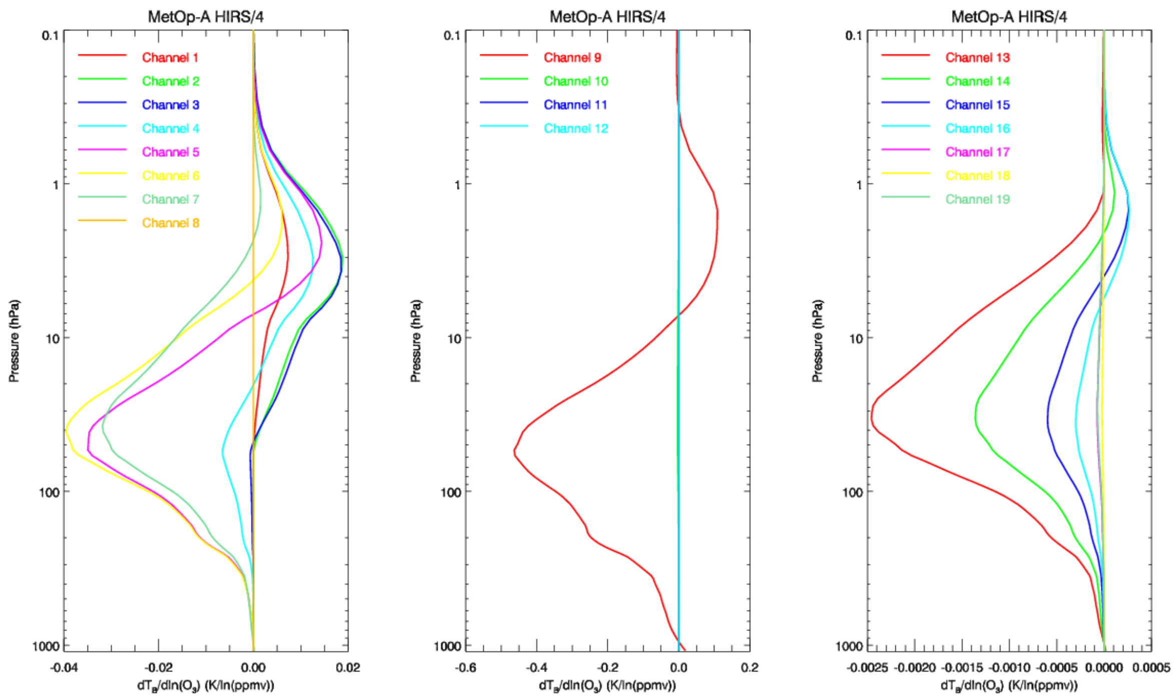

Figure 6: Ozone Jacobians for METOP-A HIRS channels.

6-14, the moisture sensitivity is small and these channels can be considered primarily temperature channels. Fig. 6 shows the Jacobians with respect to ozone for the HIRS/4 channels. Again note the changes in scale. Channel 9 shows the most sensitivity to ozone and for that reason is often referred to as the ozone channel (but still has a strong temperature signal).

As mentioned earlier, for the lower peaking channels a significant portion of the signal can come from the surface. This surface signal is influenced by both the skin temperature and the surface emissivity. Figs. 7 and 8 show the surface emissivity for a infrared channel and a microwave channel, respectively. Note the variability of the surface emissivity over land and ocean. Small errors in the emissivity (~.01) can result in differences in the simulated brightness temperatures of over 1K. So it is essential to get the emissivity estimates correct if channels with a significant surface contribution are used. The infrared emissivity is close to 1 over land and ocean, but can vary significantly due to the underlying surface characteristics and look angle. In the microwave, there is a much larger difference over land and sea with the surface emissivity over land being close to 1, while over the ocean it is close to 0.5. This large difference in surface emissivity over land and ocean can make the detection of clouds and precipitation problematic over the land and make it difficult to apply the same quality control algorithms in all situations. Also, since the size of the field of view is generally larger for the microwave instrument, it becomes very important to account for the distribution of surface types in the field of view to properly use data in transition areas (coastal, lakes). This is especially difficult when lakes are freezing or thawing in the spring and fall where the emissivity will go from around .5 for water and close to 1 for ice.

The observation error specification is another important part of the satellite assimilation process. The general observation error specification is discussed in the presentation by Desroziers in this volume. However, there are a few important points for the observation error for satellite data. The observation minus simulated observation variances for satellite data are often close to the expected instrument error variances, meaning that there is not a large signal in the observations. Because of this, the data must be used very carefully to extract as much information as possible. The observational error covariance matrix often can also be non-diagonal because of correlated errors in the radiative transfer or correlations which come from the instrument design. For that reason, the specification of the correlation can become important. However, as shown in Bormann et al., 2011, while the correct specification of the correlated error can significantly reduce the amplitude of the correction to the analysis, the incorrect specification of the correlations can amplify the correction. Thus, one must be careful specifying the correlated error for the satellite data.

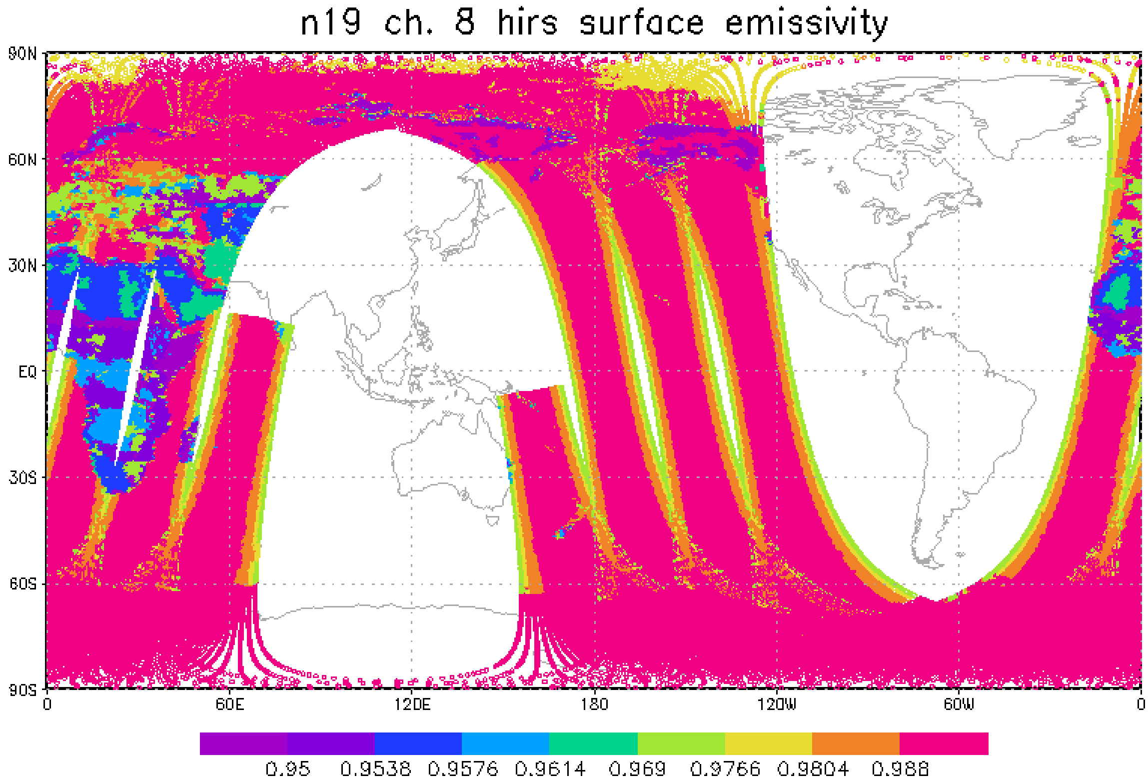

Figure 7: NOAA-19 HIRS channel 8 surface emissivity estimates.

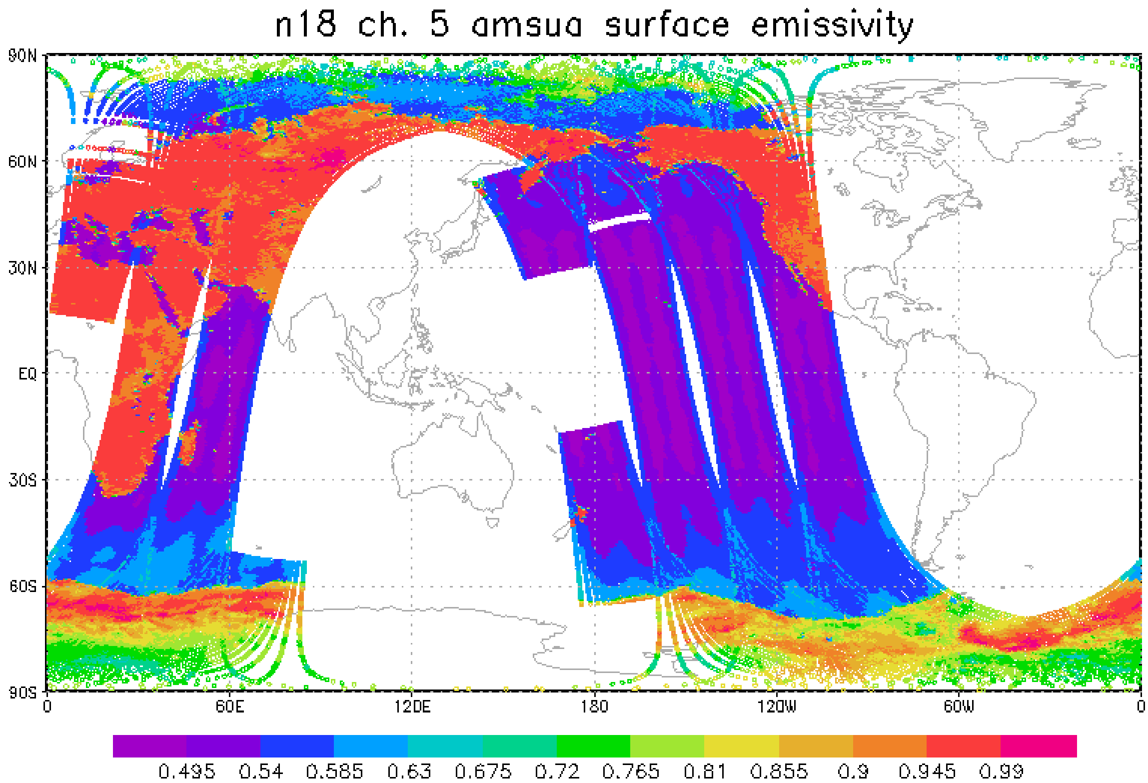

Figure 8: NOAA-18 AMSU-A channel 5 surface emissivity estimates. Note different scale to Fig. 7.

The difference between simulated and observed observations can also show significant biases (larger than the variance). These biases can come from inadequacies in the characterization of the instrument, deficiencies in the forward model, errors in the processing of the data and/or biases in the background field. This bias should be removed unless the bias comes from the background. Unfortunately it can be difficult to determine if the error is from the background or from the other sources. Currently NCEP and many other operational centres use a process for bias correcting radiances which attempts to account for a scan position dependent bias correction which has a different correction for each channel and each scan position and an air mass dependent bias correction. Fig. 9 shows an example of the scan dependent bias for 4 channels for AMSU-A on 2 different satellites. Note that the bias is not symmetric across the scan, is different for each channel and is significantly different for the different satellites with similar instruments. An air mass bias correction is the second component of the bias correction. This correction attempts to correct the state dependent component of the bias through a linear state-dependent equation. The coefficients for the state-dependent predictors are specified to reduce the air mass bias. The predictors in the linear equation can be state dependent and can include values such as an integrated lapse rate, layer mean temperatures, skin temperature, observed brightness temperatures, total precipitable water, cloud liquid water, etc.

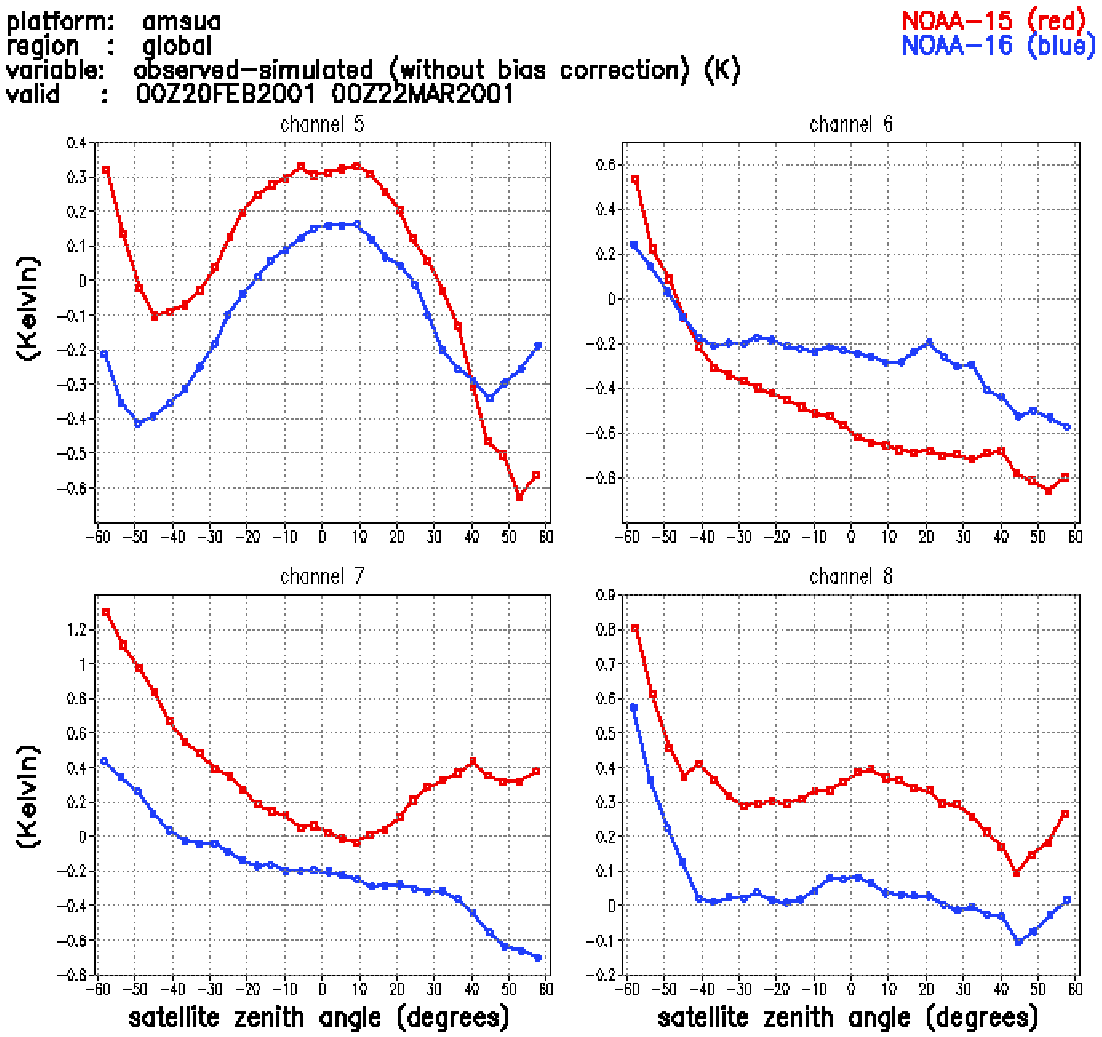

Figure 9: Cross-track average differences for NOAA-15 and NOAA-16 channels accumulated from 20 Feb. – 22 Mar. 2001.

The coefficients in the angle dependent and air mass bias correction can be specified either externally to the analysis problem or by incorporating the coefficients directly in the variational assimilation. For the second case, n (where n is the number of predictors) additional analysis variables are added for each channel to the control vector and solving for these variables along with all the other analysis variable (Derber and Wu, 1998 and Dee and Uppala, 2009). This has become known as variational bias correction. The major difficulty with this type of bias correction is its inability to distinguish between model bias and observation-guess bias when no other data is available. The inclusion of the bias correction can result in a significant reduction of the bias and variance for the observations and more a more normal distribution of the observation – background difference (see Fig. 10).

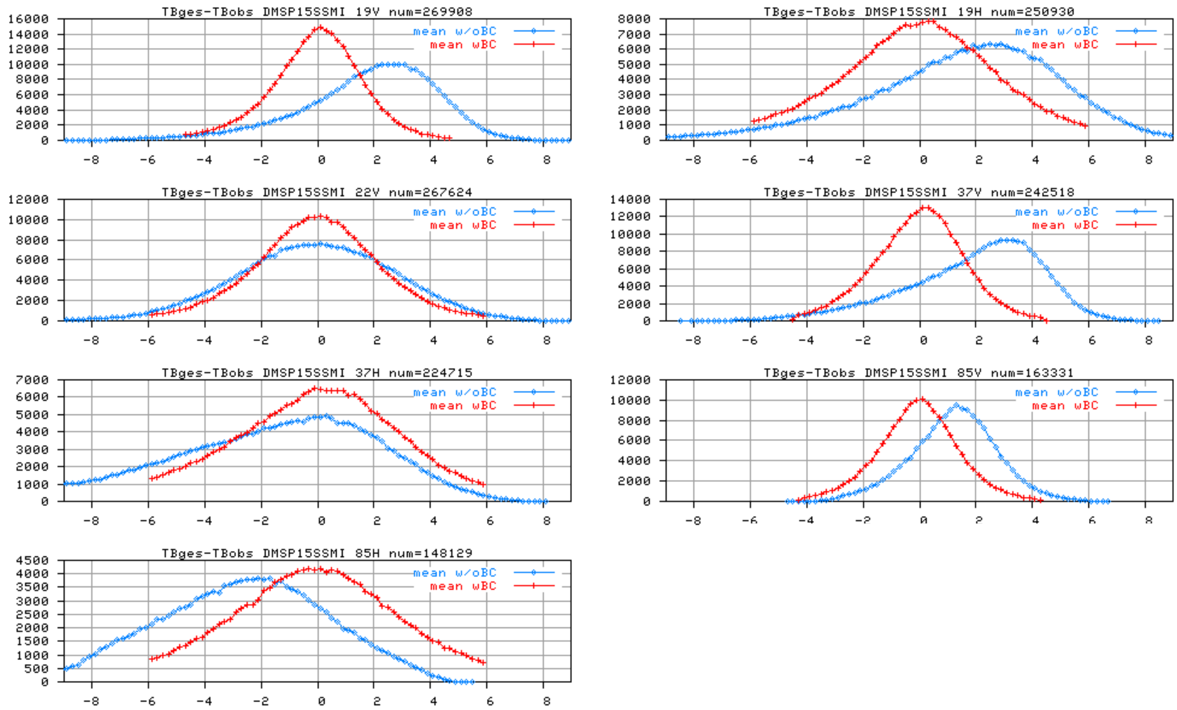

Figure 10: Histogram of observation – simulated differences with (red) and without (blue) bias correction for SSM/I observations.

The quality control step may be the most important aspect of satellite data assimilation. Data should be removed from the assimilation system (or given very low weight) if it contains gross errors or cannot be properly simulated by the forward model. For satellite data, most quality control decisions occur because of instrument problems, clouds and precipitation simulation errors, surface emissivity simulation errors, or data processing errors. For infrared radiances, the upwelling radiances are blocked by clouds and the top of the cloud becomes a new higher level surface. We can use channels with their signal completely above the top of the clouds because these will not be impacted by the cloud. However, since the channels are sensitive to deep layers of the atmosphere often only a relatively small subset of the channels are not influenced by clouds. For the microwave, the observations are much less impacted by clouds (especially thin, non-precipitating clouds) and the data can be used over a much larger area of the earth. Since it is often difficult to properly estimate the surface emissivity and skin temperature characteristics over land, snow, ice and mixed surfaces, the channels influenced by the surface are often not used in these areas.

It is essential to have a good system for monitoring the quality of the satellite data. Often the NWP centres notify each other about problems with instruments prior to being notified by the data providers. Changes in the statistics for particular channels can also result from changes in the assimilation systems. For that reason, it is useful to be able to compare monitoring sites from other operational centres. The numerical weather prediction group from the International TOVS Working Group (ITWG) has assembled links to many of the operational monitoring sites. The location of these links can be found at https://groups.ssec.wisc.edu/groups/itwg/nwp/monitoring.

There is one additional step often applied to the use of satellite data: thinning or superobbing. One of the two can be used to reduce the redundancy in the data, to reduce the effects of spatially correlated errors, and to reduce computational expense. Thinning is often applied to reduce the spatial or spectral resolution of the observations by selecting a reduced set of locations or channels. When using thinning one can use "intelligent thinning", choosing observations which are most representative or most likely to pass later quality control. Superobbing combines spatial or spectral resolution by combining locations or channels. Doing this can reduce the random noise in the data while reducing the volume of the data. The use of reconstructed radiances (used to compress the information in advanced IR sounders' spectra, Antonelli, 2004) can be considered a type of superobbing. Note that a more complex version of superobbing can be developed which can use higher moments contained in the data (Purser et al., 2010).

3. The Future

Over the last 20 years there has been tremendous progress in the use of satellite data. It is interesting to speculate where the most important developments will occur in the future. It is anticipated that the use of the observations available from satellite data will allow the assimilation of many new variables such as clouds and precipitation, trace gases and aerosols, land surface characteristic and ocean characteristics. Many of these areas were discussed further in other presentations at this seminar by Mahfouf, deRosnay, Moore, Haines and Simmons.

The inclusion of additional analysis variables results in several challenges which emphasize the integration of all components of the assimilation system. One of the most important challenges is ensuring that the proper balance between the variables (e.g., clouds, winds and temperature, ozone, skin temperature, etc.) is properly incorporated into the system so that the improvements in the initial analysis are retained for longer term forecasts. The specification of background error covariances (including cross variable correlations) will be extremely important in using the information in the data at the correct scales and making the proper adjustments to other variables. The inclusion of clouds and precipitation will result in strong nonlinearities in the minimization problem. The model bias remains an important component of data assimilation and the use of observations. If the model is biased in the production of the analysis variables, the analysis will have a very difficult time correcting the bias (especially if the variable is not well observed). In most cases, the model bias should be eliminated or reduced before assimilation can be attempted. It is important to remember that the details of an assimilation system are extremely important in determining the success of the system.

In general, the authors have noted that the satellite instrument community is very good at designing new instruments which produce volumes of observations which increase at a rate equal to or greater than the rate that they can be used (both computationally and scientifically). For that reason, much of our future effort is expended just keeping up with the new instruments as they become available e.g., COSMIC-2, NPP, GOES-R, FY3, etc.). Note that we are not now using any of the current data to its full potential. Significant future work can be directed towards improving the use of the data by enhancing the ability to simulate the skin temperature and emissivity, to properly account for all aspects of the observations geometry, to improve cloud detection techniques, to improve (or hopefully remove) bias correction and to improve thinning/superobbing techniques.

4. References

Antonelli, P., H.E. Revercomb, L.A. Sromovsky, W.L. Smith, R.O. Knuteson, D.C. Tobin, R.K. Garcia, H.B. Howell, H.-L. Huang, and F.A. Best (2004). A principal component noise filter for high spectral resolution infrared measurements, *J. Geophys. Res.*, **109**, D23102--23124.

Bormann, N., Collard, A. and Bauer, P. (2010), Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *QJRMS*, **136**: 1051–1063. doi: 10.1002/qj.615

Cucurull, L., 2010: Improvement in the use of an operational constellation of GPS radio occultation receivers in weather forecasting, *Wea. and Forecasting*., **25**, 749-767.

Dee, D. P. , UppalaS., 2009: Variational Bias correction of satellite radiance data in the ERA-Interim reanalysis, *QJRMS*, **135**, 1830-1841.

Derber, J. C. and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev*., **126**, 2287 -2299.

Desroziers, G., L. Berre, B. Chapnik and P. Poli, 2005: Diagnosis of observation background and analysis-error statistics in observation space. *QJRMS*., **131**, 3385–3396.

Healy, S., J. Eyre, M. Hamrud, and J.-N. Thepaut, 2007: Assimilating gps radio occultation measurements with two-dimensional bending angle observation operators. *QJRMS*, **133**, 1213–1227.

Purser, R.J., D.F. Parrish, M. Masutani, 2000: Meteorological observational data compression: An alternative to conventional “super-obbing”, NCEP Office Note **430**, available at: http://www.ncep.noaa.gov/officenotes/NOAA-NPM-NCEPON-0006/01408B82.pdf.