5. Evaluation of MERRA through innovation statistics

The differences between the observations and the forecast background used for the analysis (the innovations or O-F for short) and those between the observations and the final analysis (O-A) are by-products of any assimilation system and provide information about the quality of the analysis and the impact of the observations. Innovations have been traditionally used to diagnose observation, background and analysis errors at observation locations (Hollingsworth and Lönnberg 1989; Dee and da Silva 1999). At the most simplistic level, innovation variances can be used as an upper bound on background errors, which are, in turn, an upper bound on the analysis errors. With more processing (and the assumption of optimality), the O-F and O-A statistics can be used to estimate observation, background and analysis errors (Desroziers et al. 2005). They can also be used to estimate the systematic and random errors in the analysis fields. Unfortunately, such data are usually not readily available with reanalysis products. With MERRA, however, a gridded version of the observations and innovations used in the assimilation process is being made available. The dataset allows the user to conveniently perform investigations related to the observing system and to calculate error estimates. Da Silva (2011) provides an overview and analysis of these datasets for MERRA.

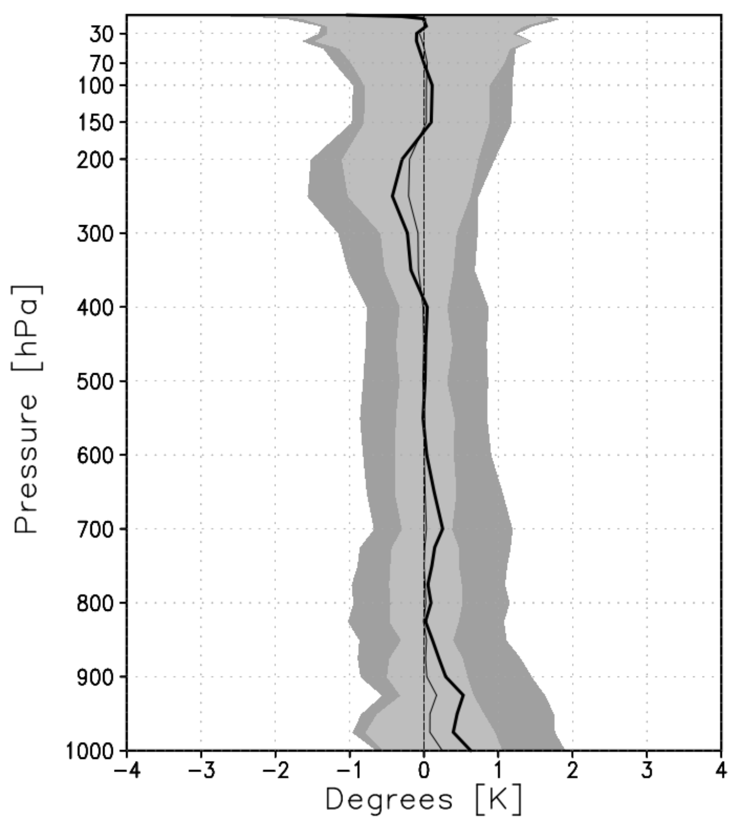

Figure 10: Vertical profile of global mean O-F (thick curve) and O-A (thin curve) residuals (K) for radiosonde temperature observations as a function of pressure level (hPa) during January 2004. The dark and light shading indicate +/-1 standard deviation from the mean O-F and O-A values, respectively.

The global mean O-F and O-A statistics for radiosonde temperature observations at different pressure levels are shown for January 2004 in Figure 10. The biases are relatively small (less than 0.5 K) at most levels, with a cold bias (positive O-F) in the PBL and a warm bias in the upper troposphere, consistent with the analysis biases against independent MLS observations discussed earlier (Figure 5). Interestingly, these O-F statistics change with time (Figure 11), especially in the upper troposphere. Since in the reanalysis the model does not change and there is no indication of degradation in the radiosonde observations themselves over time, we conclude that other observation types contribute to these changes in the agreement between the analysis (and also the background forecast) and the radiosondes. This issue is explored further in Figure 12. Even before the increase in the bias, there is a decrease in the standard deviation of the radiosonde innovations associated with the decrease in the number of radiosonde observations.

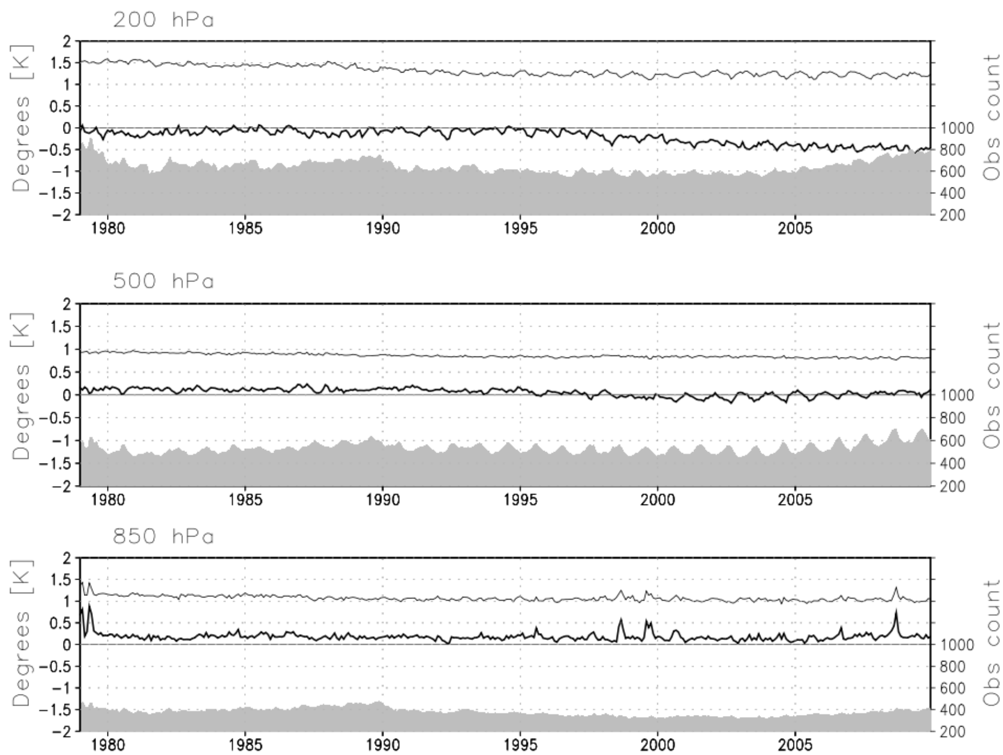

Figure 11: Time series of the monthly global mean (thick curve) and standard deviation (thin curve) of O-F residuals (K, left axis) for radiosonde temperature observations at 200 hPa (top), 500 hPa (middle) and 850 hPa (bottom). Negative mean values indicate that the observations are colder, on average, than the background. The shaded curves indicate the monthly mean data counts (right axis) for each 6-hr assimilation cycle.

Figure 12a shows the time series of monthly innovation statistics for radiosonde temperature at 300 hPa. The thin black line depicts the spatial-temporal mean O-F for each month. Comparison with the same statistic for aircraft temperatures (Figure 12b) shows that the increase in the magnitude of the upper tropospheric bias with respect to radiosondes starting in the mid- to late nineties coincides with an increase in aircraft observations, which have a warm bias (Cardinali et al. 2003, DU09). As pointed out by

DU09, after 1999 the mean analyzed temperatures are increasingly determined by the more numerous aircraft data, especially in the Northern Hemisphere, even though the observation error specified for radiosondes tends to be slightly lower than that specified for aircraft (0.65 K for radiosondes and 0.8K for most aircraft observations at 300 hPa in MERRA).

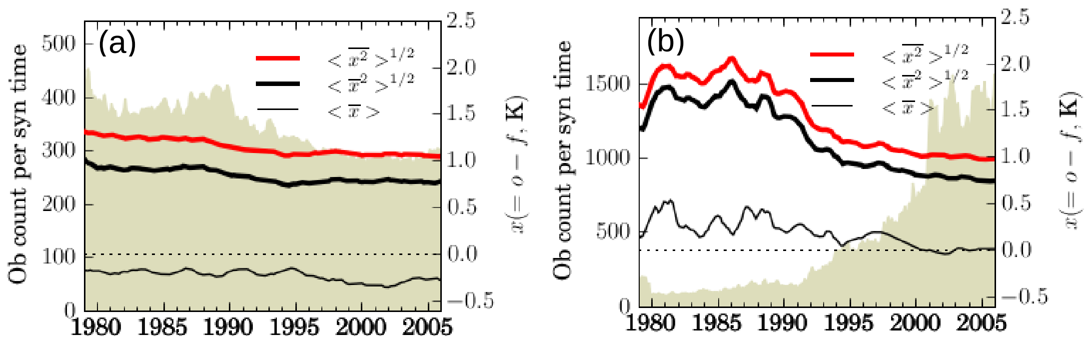

Figure 12: (a) Time series of monthly global mean O-F statistics for radiosonde temperature observations at 300 hPa. The thin black line shows the global-mean (angle brackets) monthly-mean (overbar) difference, the thick black line shows the spatial RMS of the monthly mean, and the red line shows the spatial-time RMS, all in degrees K (right axis). Shading represents the number of observations per synoptic time (left axis). Curves have been smoothed with a 12-month running mean. Panel (b) shows the same statistics for temperature observations at 300 hPa, but taken from aircraft.

Two complementary statistics are also depicted in Figure 12. The thick black line shows the spatial RMS of monthly mean values in each grid box (Da Silva, 2011). This statistic shows the contribution of the spatial variability of the O-F bias, which can be ascribed in part to the instrument inhomogeneities but most likely reflects the large-scale structure of the background bias. The red curve depicts the space-time RMS for each month. The difference between the red and thick black lines offers an indication of the contribution of synoptic scale eddies to the O-F misfit. These statistics indicate that the dominant components of the background error are systematic rather than random.

The innovations may be thought of as the correction to the background required by a given instrument, while the analysis increment (A-F) is the consolidated correction once all instruments, observation errors, and background errors have been taken into consideration. The extent to which the O-F statistics for the various instruments are similar to the A-F statistics reflects the degree of homogeneity of the observing system as a whole. Using the joint probability density function (PDF) of innovations and analysis increments, da Silva (2011) introduces the concepts of the *effective gain* (by analogy with the Kalman gain) and the *contextual bias*. In brief, the effective gain for an observation is a measure of how much the assimilation system has drawn to that type of observation, while the contextual bias is a measure of the degree of agreement between a given observation type and all other observations assimilated. For details of the computation of these quantities and their interpretation, the reader is referred to da Silva (2011).

Figure 13 shows a time series of the contextual bias and effective gain, along with the correlation between O-F and A-F, for radiosonde virtual temperature at 300 hPa. Consistently over the record, about 60% of the innovations are realized as analysis increments, with the O-F and A-F being correlated at 90%. These observations exhibit a larger (more negative) contextual bias in the 1980s and early 1990s, probably due to the influence of data from thermal infrared sensors such as the High resolution Infrared Radiation Sounder (HIRS). This bias decreases more or less steadily with time after 1990, probably due to the influence of higher-quality satellite data available in recent years. Note that this contextual bias reflects the observations themselves, not the bias of the background (as seen in Figure 12), which is influenced by the weights given individual observations in a prior analysis. The fact that the contextual bias of the radiosonde observations at 300 hPa decreases over time is another indication that the increase in background bias relative to radiosondes seen in Figure 12 is due to the influence of other observations (the aircraft data) on the analysis, as inferred above.

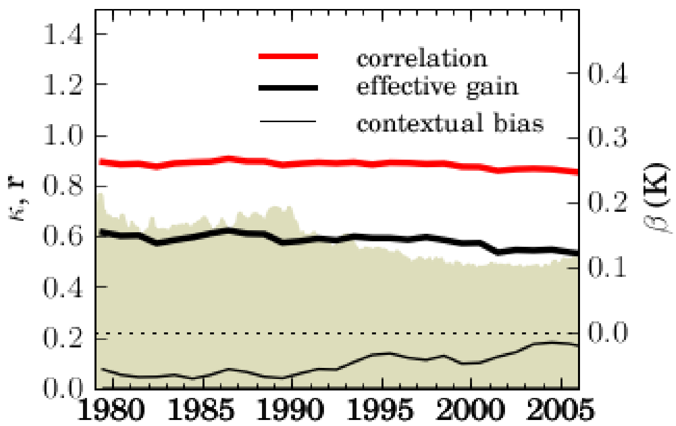

Figure 13: Time series of correlation (red), effective gain (thick black) and contextual bias (thin black) based on radiosonde virtual temperature data at 300 hPa. The axis for correlation and effective gain is on the left; the axis for the contextual bias is on the right.

With MERRA's gridded observation and innovation data sets, a wealth of information is available for examination of the quality of the analyses and how the different observations impact the analyses and interact with each other. Such examinations can be conducted regionally or globally and should provide useful information for the next generation of reanalyses.

6. Climate Variability

Many aspects of the quality of MERRA products are presented in other papers mentioned in the Introduction. In the next three sections, we touch on just a few fields that highlight improvements over earlier-generation reanalyses and on some of the issues that will still need to be addressed in the next generation.