Observational requirements for future reanalyses

Sakari Uppala

ECMWF, Shinfield Park, Reading

1. Introduction

Current practice is for each reanalysis project to take responsibility for collecting historical observations as best as it can and merging the various observations into an input dataset for the analysis. This rather ad hoc approach to the collection and merging of data has worked well, but it often relied on the good will and personal devotion of a few key people in the responsible organizations.

As we begin work on the next generation of reanalysis products, it has become clear that it is not sufficient to simply use the observational input from a previous reanalysis, since many shortcomings in the data and their usage are usually identified in the evaluation of the earlier reanalysis products. Each successive reanalysis project can, in fact, extract a greater number of observations and more accurate information from them by taking advantage of new developments in data-assimilation systems, the handling of biases in observations, and in the metadata.

NASA, NOAA, NSF sponsored Workshop on “The Development of Improved Observational Data Sets for Reanalysis: Lessons learned and Future Directions” was held in September 2005 at University of MD Conference Centre. In the following general points from the workshop are summarized and the main recommendations are listed in the end.

2. Current practice in reanalyses

The common practice has been to create merged input datasets: one input stream for all the conventional observations, and separate input streams for each satellite system. This is a major task that requires continuous effort, and broad knowledge from data experts on the characteristics of the observing systems, including their history. New data continue to become available from both data archaeology and the current suite of operational observing systems. The result is that the magnitude of this effort increases each year, since the massive amount of new operational data has to be merged into the input streams. The need to merge data from several sources is illustrated in Figure 1 for Aircraft data.

Reanalysis makes use of the input data and utilizes all the available metadata. At the same time reanalyses create the feedback metadata for each input stream. This metadata contains the information about how the reanalysis has used each datum with the departure information attached, schematically shown in Figure 2.

Future reanalysis projects are dependent on the activities of the data collection centres around the world. Since in the view of the workshop, reanalysis is an on-going iterative process, a more coordinated program is required to more efficiently update and enhance the input data streams. This includes taking advantage of the feedback from the previous reanalysis efforts in the creation of the next input dataset, see Figure 3.

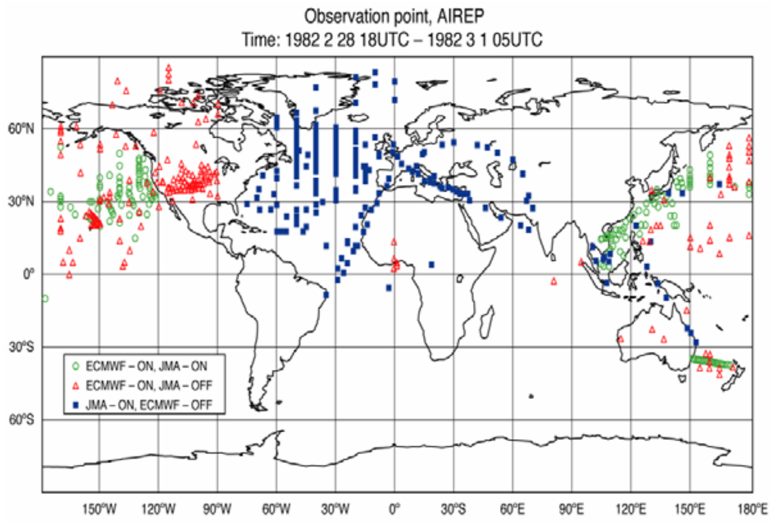

Figure 1 The importance of the data merge between the main centres is highlighted by a comparison of the holdings of Aircraft data in the JMA and ECMWF archives for a two day period. Green circles indicate data found in both archives, red triangles indicate data found only in ECMWF archive and blue squares indicate data found only in JMA archive.

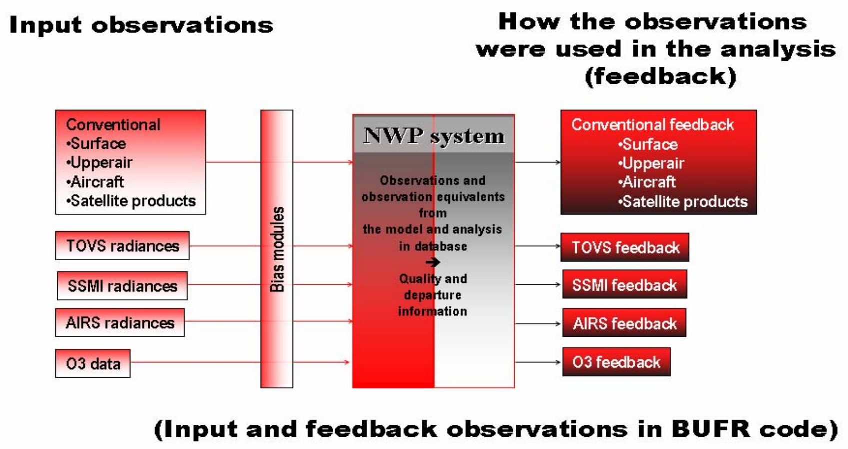

Figure 2 Dataflow through a data assimilation system. A value added product "feedback file" is created for each input stream.

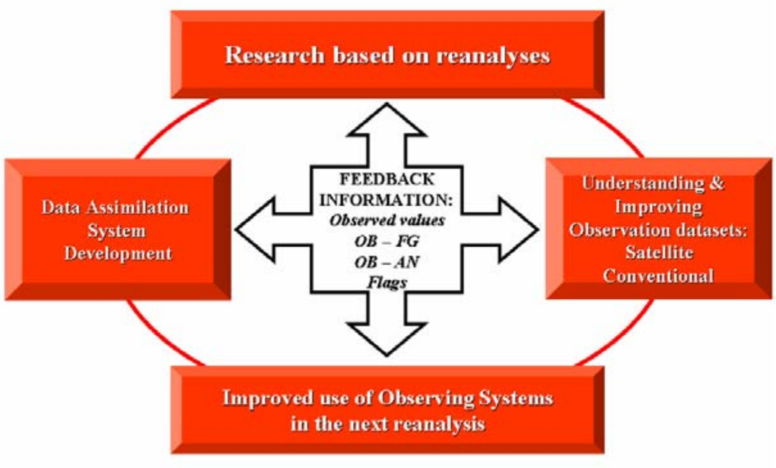

Figure 3 A schematic presentation of the interaction between research, data preparation and the next reanalysis.

The creation of input datasets for reanalysis can be viewed as an important contribution to our stewardship of global climate observations, their maintenance for future use, and the extraction of reliable information concerning climate variability and change. It is a unified approach to the observing system and its history spanning both conventional and satellite data. It is a major effort that needs on-going institutional support. It also needs international support to open data policies so that access to important data can be achieved worldwide.

Recent technological advances have been important not only in terms of improved data storage and data handling, but also in the availability of high performance data base systems. These systems allow the maintenance of different data sources separately using version control, so that reanalysis projects can extract or enquire new merged versions from the sources. Furthermore, the databases can be appended by new metadata and information from the feedback records of different reanalyses. This would facilitate studies based on observations alone, and should help to strengthen the ties between observationally based research and research based on reanalyses.

3. Emerging capabilities

The last few years have seen the development of a number of promising new techniques for identifying and correcting observational bias and other errors. In particular, bias correction techniques have been developed for radiance data.

The procedure, known as radiance bias correction, in fact has a long history in the quantitative use of satellite data in both retrievals and numerical weather prediction assimilation schemes. Radiance bias correction allows for correction of unresolved errors in a number of areas of the satellite sounding and inversion. These sources include forward radiative transfer errors, in situ validation instrument errors, errors in instrument end-to-end calibration, and errors in the numerical weather prediction models base climate state. Reduction of each of these sources of error is an ongoing effort by many research groups.

There are also extensive efforts aimed at the calibration of satellite instruments. The calibration and inter-calibration of radiance data for climate studies can be considered as a three-step process. The first step, referred to as the nominal calibration, involves using the best procedures to provide for the calibration of a single instrument on a single satellite. This step is also common to reanalysis for they also require the best absolute calibration of each instrument on each satellite. The second step is referred to as the normalized calibration. The normalization is usually accomplished by identifying one instrument as the baseline and then adjusting the systematic biases of the other instruments in the series to that baseline instrument. A final step involves adjusting the normalized calibration to some ‘absolute’ geophysical observation of the variable that is being retrieved, and is hence referred to as the absolute calibration.

As mentioned earlier there is also work being done to better take advantage of reanalysis feedback files, Figures 2 and 3. The possibility to use this information to improve the quality control of the observational data sets in an iterative manner has been shown for the radiosonde data and for some of the radiance data. However, it is not that simple because the first guess background fields are not observations but a blend of model, and it’s assumptions and simplifications of atmospheric processes, and observations. Nevertheless, there is a significant potential for using the reanalysis feedback files, for example, when observations and first guess background fields systematically differ over extended regions or time there is reason to investigate the observational data sources for possible errors. Gleaning and applying this additional quality control information is currently under utilized and is a new area of effort for the data stewards.

4. Vision for the future

In order to optimize the use of observational data for both reprocessing and reanalysis, the workshop promoted the adoption of the concept of scientific data stewardship. Scientific data stewardship (SDS) is the new paradigm in data management consisting of an integrated suite of functions to preserve and exploit the full scientific value of environmental data. These functions are the careful monitoring of observing system performance for long-term applications, the generation of authoritative long-term records, for both reprocessing and reanalysis, from multiple observing platforms, the assessment of the state of the atmospheric, oceanic, land, cryospheric and space environments, and the proper archival of and timely access to data and metadata.

To promote full exploitation of the scientific value by current and future users, four functions, each with several constituent components, must be achieved. The first function is to provide real-time monitoring of the observing system performance for long-term applications. Such monitoring requires the establishment of tracking tools necessary for the detection of changes in the observing system as well as in the observation record. One example is the detection of small biases in the instrumental record. These biases can then be minimized or eliminated through efficient coordination with network operators.

The second function is generating authoritative, long-term records. This function will preserve and enhance the value of the irreplaceable historical data by conducting rigorous data analysis and research to validate and improve these authoritative records, and by reanalysis and reprocessing and enabling others to participate in these. For reanalysis, the primary techniques involve 3- and 4-dimensional variational analysis using a model to fuse together data from disparate observing systems like direct measurements from ground-based networks and indirect measurements from remote sensing instruments on satellites.

The third function uses the authoritative records to assess the current state of the environment and to put it in historical perspective. Long-term trends on local, regional or global scales can be determined and estimated for the future. In addition the authoritative records can be used to detect changes in environmental conditions between different time periods and different environmental regimes. The reanalysis framework provides a comprehensive framework for assessing the impact of different observing systems on the end product. At one end of the spectrum of reanalysis approaches all the available observations are used to obtain the best estimate of the earth system at all times. This type of reanalysis can be highly useful for studies of the complex interactions between, for example, the biological and geophysical processes. At the other end of the reanalysis spectrum, use of a consistent set of observations for different observing system epochs has led to a strategy of differing analysis time periods.

The final function, insuring complete archival and access capabilities, requires that metadata, direct observations, and fundamental records from satellite and in situ platforms be comprehensive, complete and preserved, in perpetuity. Open, efficient access to the metadata, products, and data streams must be insured, and data made available in useful formats. The metadata bases become particularly important as the archives are cyclically refined. It is here that improved understanding of observations is captured, e.g. physics of the instrument design, basic and reanalysis-determined quality control, and lineage of use.

5. Recommendations

The workshop gave the following recommendations:

The key recommendation of the workshop

WCRP Observations and Assimilation Panel (WOAP) to appoint a working group of experts charged with developing a plan for "The On-going Development of Improved Observational Data Sets for

Reanalysis", that describes the necessary resources, infrastructure, institutional commitments, and coordination on technical issues outlined in this report.

Action: Get commitments from member countries to carry out specific tasks as outlined below in the scientific and technical recommendations and action items.

The workshop also identified the need to raise the profile of activities concerning the pre, post, and bias processing of observations and the related science. It was recognized that the improved understanding of climate variability and change depends more than ever on the quality of the historical observations and their effective use in reanalyses.

The meeting highlighted that recovery and updating of observational data is an ongoing effort and reanalyses themselves are part of this activity. A new reanalysis project needs to get the latest versions of data from various sources, and a single operational archive on its own does not provide enough information. Keeping track of the global observational dataset is crucial for the success of future reanalysis efforts. This task is increasing in complexity with many more satellite sources being available in recent years and through rescue and improvement of conventional data.

There was general agreement that one of the most pressing needs was for the various data centers to conduct a full international coordinated inventory of their conventional and satellite data holdings. The scope of the inventory would be global and include all observations relevant to reanalysis of the Earth System encompassing observations of the atmosphere, ocean, land, and cryosphere.

The specific technical and scientific recommendations were as follows:

Recommendation 1:

All main centres should prepare inventories of reanalysis observations on the level of observation records

Recommendation 2:

A collaboration be formed that can sustain a data refresh cycle and create high quality merged datasets for reanalyses.

Recommendation 3:

Develop improved record tracking control for observations to further improve the use of feedback data from reanalyses targeted especially for data providers/developers.

Recommendation 4:

The observational, reanalysis, and climate communities should take a coordinated approach to further optimizing reanalysis for climate.