The US Reanalysis Program: Climatology for the New Millennium

1. Introduction

In the last decade a new methodology has completely changed the traditional approach to climatology. A state-of-the-art data assimilation system is used to *reprocess all past environmental observations*, combining them with *short forecasts* in order to derive the *best estimate of the state and evolution of the environment*. Since the statistical combination of the forecast and observations is denoted in operational applications as analysis , the new method is usually known as ***Reanalysis.*** Perhaps a better name would be a ***Reintegration*** *of our sum total knowledge about the environment into an easily accessible data base*.

Because it combines observations with forecasts using current models and advanced methods that are kept unchanged, reanalysis provides, for the first time, an optimal estimate of the 4-dimensional state of the atmosphere, free from abrupt changes due to changes in the assimilation system. By using state-of-the-art models, the reanalyses also provide an invaluable history of our best estimates of a host of difficult to observe quantities such as latent heating, surface fluxes and vertical motion. Modern technology offers no simple way to routinely observe these critically important parameters for daily and climate prediction.

Three major global atmospheric reanalyses have been already carried out:

1) The NCEP/NCAR 50-year Reanalysis 1, covering from 1948 to the present, followed by the NCEP/DoE 20-year Reanalysis 2.
2) The ECMWF 15-year Reanalysis, covering 1979-1993.
3) The NASA/DAO 17-year Reanalysis, covering 1980-1996

In addition, the University of Maryland has performed a preliminary 50-year reanalysis of the oceans.

A modern reanalysis is to traditional climatology as a modern car is to a model T-Ford or even a 1960's Chevy. Because of the continued improvements in technology, it makes possible advances that were impossible 50 or even 10 or 20 years ago. Because the output of the reanalysis is given on a uniform grid, without missing data, anybody, from a new graduate student, through a veteran researcher to an operational weather and climate forecaster, can now carry out in a few weeks an assessment of climate properties that would have taken much longer to make before its advent. As a result, there has been a huge amount of research in recent years that would have been impossible before, increasing very much our understanding of the evolution of the atmosphere. It is estimated that 5-15 thousand papers and studies have been carried out just in the last few years using the reanalyses, and their use is growing exponentially. Operational weather and climate predictions have benefited from the research by getting greatly improved models and better data for verification.

However, like cars or computers, reanalyses should improve as the science and technology of data assimilation improves with time and from competition between different approaches. Each of the three pioneering efforts mentioned above had significant problems, and it has already been found very useful to have several reanalyses in order to check the validity of each result.

Reanalysis combines the skills of modern Meteorology, Oceanography, Hydrology, Remote Sensing and Computer Science to create environmental databases from which systematic changes can be better assessed. A modern reanalysis program will allow this database to be continually improved as we learn more about the physical, chemical and/or biological processes that control our environment. As such it both serves and benefits from science. It serves science by identifying gaps in our knowledge and by transitioning progress to users. It benefits from science by incorporating research results as they become available. Because atmospheric, hydrological and ocean models are becoming more realistic, and the availability of faster supercomputers allow higher resolution, reanalyses, like the operational forecasts on which they are generally based, should continue to improve. Moreover, a new demand for estimates of the evolution of additional physical, chemical and biological processes is likely to appear as the technology evolves and additional uses for the reanalysis become apparent.

Because of the advances mentioned above, the main conclusion of the Workshop on Global Reanalysis that took place on June 5 and 6 2000 that long-term ***Global Reanalyses should be carried out every ten years***, when the evolution of the technology justifies it. This time scale is typical of major improvements in models and data assimilation, which make previous reanalyses obsolete. It was furthermore concluded that they should be ***Earth System Reanalyses*** rather than just global **atmospheric** reanalyses, and include ocean and land components as an integral part. New products may become necessary as new applications (e.g., air chemistry, health) are developed.

It is important to utilize both national and international resources for this effort. For example, if several agencies such as NOAA, NASA, NSF, DoE, Navy and others, as well as university resources, carry out a periodic global reanalysis every 10 years, it could alternate with a similar effort in Europe, resulting in the availability of a state-of-the-art reanalysis every 5 years. The preparation for a major reanalysis takes several years, as do its execution, posterior distribution to the community and evaluation; therefore pacing these efforts on a decadal time scale is appropriate. But it is not possible to start each of these efforts from scratch: *S*uch effort should remain the main mission of a dedicated core of scientists.

Global reanalyses are used for climate monitoring, climate prediction, studies of the international (WCRP) CLIVAR program, as well for applications such as prediction and monitoring of climate related health problems, stratospheric transports and chemistry, boundary conditions for regional models, and many others. In fact, reanalysis (combined with a current Climate Data Assimilation System or CDAS) should be the backbone of a climate services program. The first reanalyses had problems that made them unusable for some of these applications. For example, spurious changes in the perceived climate appeared in the reanalysis at the times when new observing systems were introduced, such as satellites in the late 1970 s. These changes in the observing system have continued and, in fact, the World Climate Research Program has identified the need to solve the problem of these spurious changes and trends as the most important problem in reanalysis. The coarse resolution of the initial reanalyses made them inappropriate for many regional applications (e.g., health related problems). The problems of the models in estimating quantities such as precipitation, surface stress and turbulent fluxes of heat and momentum have made very difficult the applications that require coupling to the oceans. The assimilation over land (known as Land Data Assimilation System or LDAS) is in its infancy. All these problems are slowly but surely being attacked in the modeling community, but unless the advances are used in a new reanalysis, their benefits will not reach the general public.

Following this introduction, the report contains a series of sections elaborating on why we need a reanalysis program (as opposed to simply another reanalysis project). They include the expected benefits (Section 2), a description of current and potential uses, including climate and weather prediction, hydrology, oceanography, health, and regional reanalyses (Section 3), and a proposed program including components and required resources (Section 4) The expected resources are approximately $10 million/year, with about half as new funding and the remainder from existing activities incorporated under this umbrella.

2. Why a Reanalysis Program?

Rationale

A key message from the workshop was the need to embrace the idea of reanalysis as a continuing program of technology and data integration. Such a program would look across agencies to focus resources on improving how we organize, unify and extend the broad spectrum of disparate environmental observations. As we enter the new millennium, our ability to understand climate variability will be determined by how well we are able to apply new technologies to better integrate our spatially limited historical observations with our everimproving modern satellite-based global data. For example, we know that the United States tends to experience recurring extended periods of severe drought conditions, but we do not understand why this happens. This is despite the fact that there are now global satellite observations of the land surface, and we are developing sophisticated new interactive land surface models that are able to simulate the complicated interactions between the atmosphere and land (e.g. Koster and Suarez 1996). What is missing is the infrastructure to easily and consistently apply this new technology to reinterpret and extend the limited historical database of these events, and place them in the context of current climate observations.

By institutionalizing reanalysis we provide a program that looks across all data sources, across all geographical regions, across our entire historical data record, and helps to focus and develop key technologies that will allow the periodic "reintegration" of our climate data record. As the above example suggests, the need for these periodic ***reintegrations*** of our climate data is grounded in the idea that their scientific value can be greatly increased by the application of new and consistent technologies (improved models, analysis techniques, quality control, etc.). This idea is not new and, in fact, was highlighted in a National Research Council Report (NRC 1991) that outlined a strategy for a nationally focused program on data assimilation for the earth system. One of the key recommendations of that report was that "A coordinated national program should be implemented and funded to develop consistent, long term assimilated data sets (extended back to 1950) for the study of climate and global change".

Benefits and linkages

The benefits of a program of reanalysis cut across all Earth system sciences and programs. In fact, many programs now rely explicitly on the future availability of state-of-the-art reanalysis data. Linkages to research programs provide a mechanism for developing community support as well as open access to fertile grounds for new scientific findings based on the reanalysis program products.

For example, the US Climate Variability and Predictability (CLIVAR) program is an ambitious effort to identify, understand, and evaluate the predictability of the major patterns of climate variability on seasonal and longer time scales. It also aims to expand our capacity in climate predictability; to enhance the reliability of projections of climate change due to human activity; and to better understand abrupt climate changes. Research in these areas requires consistent data sets in order to understand the coupled behavior of the atmosphere, ocean and land surface, and to validate the complex models used to make predictions of future behavior. *Reanalysis products will be essential for all of these goals*. On-going efforts to improve seasonal-to-interannual forecasts and the development of suitable forecast applications can benefit greatly from long term consistent data sets for initial conditions, forcing, and validation of coupled models. These efforts include the seasonal to interannual forecasting program at NOAA, the NASA Seasonal to Interannual Prediction Project (NSIPP), and the International Research Institute (IRI). Other research efforts to study the cycles of carbon, as well as coastal research (IGBP and JGOFS) also provide a strong motivation for a reanalysis program. The weather prediction community values reanalysis products for providing the capability to undertake retrospective studies of hurricanes, major blocking events, typhoons, droughts, floods, and any number of classic systems that produced severe weather (hail, tornadoes, severe winds, heat waves, etc).

The Global Energy and Water Cycle Experiment (GEWEX) requires the data from reanalyses in order to understand the behavior of water in the earth system, and to develop quantitative models capable of predicting its variability. Currently, the WCRP is proposing the Coordinated Enhanced Observing Period (CEOP) as an umbrella for a joint GEWEX/CLIVAR project. The idea of CEOP is to use simultaneous satellite observations to link the various regional climate anomalies as they occur around the world, say, during an El Ni o event, or other times when heat sources or sinks are substantially changed. By incorporating the satellite data into a reanalysis data set we can provide estimates of the physical processes driving the variability and place the current events in the context of previous climate variations. This will allow us to better assess the causes of regional anomalies, predict their future occurrence, and possibly minimize their impacts.

The Global Climate Observing System (GCOS) program, through its Atmospheric and Oceanic Observations for Climate, has defined requirements that also can only be satisfied by globally complete gridded fields of atmospheric dynamic and thermodynamic parameters.

The influence of climate variability and change on ecosystems is receiving increasing attention. Temperature, for instance, is one of the primary factors in determining the large-scale distribution of fish and shellfish. Changes in SST and winds associated with changes in the major modes of atmospheric variability over the northern oceans have been linked to variations in the production of zooplankton, as well as to fluctuations in several of the most important fish stocks. Over land, climate variability has been associated with significant changes in plant phenology, as well as the growth, reproduction and demography of many animals. All of this includes not only longer-term changes associated with interdecadal variability, but also interannual signals, that can best be analyzed with long term consistent global gridded data sets, i.e., with reanalyses.

The carriers of human illness respond to climate change when the ecosystem in which they live is affected by that change. For instance, the outbreak of the deadly Hantavirus in 1993 in the southwestern US was caused by a period of drought succeeded by heavy rains. The drought induced a decline in many animal populations, but deer mice, carriers of the virus, recovered far more quickly than their predators. We are just at the beginning of understanding these connections and the possible consequences of climate change on ecosystems. A global understanding of the links between climate variability and ecological processes can be regarded as a new challenge, but one that can be tackled with improved retrospective, high resolution, global analyses.

National and international efforts to monitor and forecast climate variability and change, such as the proposed NOAA Climate Observations and Services Program and the Climate Information and Prediction Services project, will rely extensively on the products of repeated reanalyses and the associated current climate analyses to achieve their missions. The monitoring of climate variability requires both accurate analyses of current conditions, and a consistent historical database that enables evaluations of departures from past conditions to be evaluated. The development of better forecasting models depends upon the availability of consistent historical records in order to evaluate model performance in controlled testing, and the proper application of forecasts and other climate information products must be developed through the comparison of past events of interest with simultaneous climatic events. None of these activities will be possible without the products of a continuing series of reanalyses and consistent current climate analyses.

3. Current Use of Data from Reanalysis

Scientists from all over the world have made extensive use of data from the reanalysis projects. Many studies compare reanalysis output data with various types of observational estimates and with other reanalysis projects. Real world estimates (for example, precipitation from ground measurements and radiation from satellites) can be used to assess the accuracy of the reanalyses. Comparison of different reanalyses reveals which fields the reanalyses agree well on and which fields the reanalyses differ on, giving insight into the accuracy of different fields from reanalyses.

The evaluations typically show that the new analyses of variables such as upper air temperature, humidity and winds are much better than the same fields from operational analyses made 40, or even 10 years ago. This is due to better data assimilation methods, more complete collections of observations, and the use of consistent analysis methods over time (although the observing system has changed). When the general utility of the reanalysis data has been established, it can be used for many other purposes.

People currently use reanalysis data for research and for many applications. Table 1 is a list of many of these uses. Some of them are discussed in greater detail below. The data are especially interesting to users if several decades of output are available (as with the data for 1948 — present from the NCEP/NCAR long reanalysis). This makes possible for the first time to study longer time scales of climate variability, and if they are extended into real time, as in NCEP s Climate Data Assimilation System (CDAS), allow improved monitoring of the present climate.

TABLE 1. Some of the Current Applications of Reanalysis

a. Climate monitoring
b. Seasonal climate prediction
c. Climate variability studies
d. Boundary conditions for continental or local mesoscale models
e. Surface fluxes for ocean modeling (surface wind, radiation, sensible heat flux, precipitation, and evaporation)
f. Hydrological studies and inputs for land data assimilation systems
g. Diagnostic studies (e.g., atmospheric angular momentum, diabatic heating, atmospheric tides, teleconnections)
h. Stratospheric studies
i. Regional studies (e.g., polar regions, monsoons)
j. Nesting regional models into the global reanalysis (e.g., Project to Intercompare Regional Climate Systems, PIRCS)
k. Verification of climate models
l. Improvements to operational data assimilation and forecast models
m. Availability of data sets of observations

Operational weather forecasts

Reanalysis has improved operational forecasts, revealing errors in the use of observations, land and ocean surface parameterizations, oceanic albedo, radiation parameterizations, cloudiness and many other aspects. Preparations for a reanalysis involve extensive development work, based on examinations of previous model output by the research community, to accelerate improvement in the analysis and forecast models used for operational forecasts. Reanalysis efforts and evaluations have devoted considerable time to examining estimates of atmospheric physics such as surface fluxes and atmospheric diabatic heating that are usually not of immediate concern to operational forecasting.

A reanalysis effort produces a greater variety of output diagnostics for a much longer period of time and provides a much larger sample of the results of a data assimilation system than does an operational forecasting system. It often takes years to process results of observational efforts and to obtain the best estimates. Such estimates can be easily compared with a reanalysis that uses a modern data assimilation system to process past years of observations, but are more difficult to compare with operational analyses. The output fields from reanalysis are examined by a large number of scientists from different disciplines and compared to a large variety of independent estimates over extended periods and different periods. Collectively, these scientists have more and broader knowledge and broader and different perspectives than can be found in any operational analysis/forecast system development group. Reanalysis should allow for identification of key periods on which to test future developments of the forecasting system, and the development of metrics of their performance under a wide variety of conditions.

Seasonal forecasts

To develop numerical models for making seasonal forecasts, a long consistent set of analyses from reanalysis is needed to verify experimental seasonal forecasts on past seasons. Studies based on reanalyses offer insight into phenomena that are predictable on seasonal time scales. More empirical methods of seasonal forecasting need the best available analyses of the atmosphere and ocean over many decades as an adequate base to develop forecasting techniques based on statistics; reanalysis provides a consistent database for such techniques. The Ethiopian and Malaysian weather services are currently using the NCEP/NCAR reanalysis to study how ENSO has affected their weather in the past as guidance for the impact of future ENSO episodes. Similar efforts are taking place in South and Central America and in other Asian and African countries.

Climate monitoring

The combination of an extensive reanalysis with a real-time, ongoing climate data assimilation system (CDAS) using the same data assimilation system as the reanalysis provides a real-time monitor of the climate and estimates of current anomalies in climate not affected by changes in the data assimilation system. Current climate anomalies can therefore be readily compared to past climate anomalies, something that was impossible until the first NCEP/NCAR reanalysis was carried out. Such a system also provides current, on-going global estimates of atmospheric physics such as surface fluxes that can provide more complete understanding of climate anomalies.

Climate variability and change

The world is struggling to obtain a better understanding of climate variability, and to develop a more precise outlook for future changes in the climate. Several types of data are used for these studies. Reanalysis output data, the direct use of observations, and the use of paleoclimate data are important inputs to study climate variability. The international program for climate variability (CLIVAR) is very interested in access to reanalysis data. Variability in the climate models can be compared with atmospheric variability in reanalysis to determine if climate models produce reasonable magnitudes and patterns of variability.

Output from at least two independent state-of-the-art reanalysis projects can help evaluate them; areas of agreement and disagreement gives some indication of the accuracy of model fields and of long-term trends in the reanalyses. Such comparison revealed major problems in obtaining global long-term trends in temperature from the first reanalyses, although the NCEP/NCAR reanalysis succeeded in reproducing local long-term trends in surface temperature accurately. The next reanalyses will be designed with an objective to yield more reliable long-term trends.

Reanalyses should provide the most complete estimates for the longest period of atmospheric physics (surface and top of the atmosphere fluxes and atmospheric diabatic heating), facilitating a comprehensive understanding of climate variability and change. Examination of the first reanalyses has revealed many problems in these fields that need to be corrected in future reanalyses.

Stratospheric studies

The long-term radiosonde record potentially allows realistic analyses of the lower to middle stratosphere extending back to the International Geophysical Year, at least in the Northern Hemisphere. Reanalysis can therefore be used to determine the climatological structure and interannual variations in stratospheric flow, with quantitative diagnostics of transient events (sudden warmings) and the low-frequency (intra-seasonal and interannual) variability of high latitudes. Similarly, there should be enough data to constrain the tropical stratospheric winds, allowing quantitative examination of the long-term variability of the quasi- biennial oscillation, which is thought to have impacts on the climatology of the extratropical stratosphere and troposphere.

Transport of trace gases across the tropopause is an important component of studies of stratospheric chemistry. One process that can potentially be analyzed using conventional reanalysis (with no data extending into the stratopause region), is the complex stratosphere-troposphere exchange. With sufficient vertical resolution in the tropopause region, quantitative studies of quasi-horizontal exchange through tropopause breaks and tropopause folding events should be possible. Similarly, the upward transport of air through the tropical tropopause should be possible. One topic of extreme interest is the water vapor budget of the lower stratosphere, which depends on getting both the transport and the temperature correct.

For studies of ozone and other trace gases, quantitative results could be attained with an assimilation system extending into the mesosphere, with observational constraints up to the stratopause region. This high-top system would facilitate off-line studies of seasonal and interannual variations in trace gas transport and the ozone balance of the middle atmosphere and allow quantitative investigation of the causes of interannual variability in ozone. A high-top system would also allow for the assimilation of ozone and hence improve estimates of the radiative forcing of climate. Inclusion of the middle atmosphere in the post-1979 period should allow better understanding of the role of ozone in climate and help chemical modelers in their efforts to predict the long-term evolution of ozone until the 2050 time scale.

Modeling the oceans and the land surface

Oceanographers need ocean surface forcings (such as surface wind, surface radiation, sensible heat flux, precipitation and evaporation) from atmospheric models in order to drive ocean models. The first reanalyses provided the first global estimates of such fields for extended periods. Problems were evident in the reanalysis estimates of precipitation and surface radiation; other fields such as surface stress and sensible and latent heat flux often agreed well with independent estimates based on ship observations in regions where the independent estimates were most reliable. The accuracy requirements for driving ocean models are strict, but the oceanographers can adjust for small biases if the reanalysis output is reasonably close to reality. Many studies critically examined air-sea fluxes in the first reanalyses and suggested ways to improve them, and some have already been implemented in operational analysis/forecast systems.

The first reanalyses produced analyses every 6 hours. The land surface community believes fields are needed every 3 hours to model the interactions of the atmosphere with the land surface. Eventually the reanalyses hydrological cycles should include river flow models.

Use of global reanalysis in high resolution regional models

Mesoscale models require boundary conditions of winds, temperature and other fields; use of reanalysis output for boundary conditions has become common. Examples are models run for regions in the United States and many other countries, and to generate a climatology for Kosovo in Central Europe (for use during 1999). The NCEP regional reanalysis for North America, currently underway, uses fields from the NCEP/NCAR reanalysis as boundary conditions for the NCEP mesoscale analysis/forecast system. In the Project to Intercompare Regional Climate Systems (PIRCS), many regional climate models nested into the NCEP Reanalysis were compared. This approach has now been adopted as a powerful tool for regional climate impact studies.

Impact of reanalysis on the availability of data sets of observations

NCAR and the reanalysis centers devoted considerable careful effort to collecting and processing data sets of observations to be used in the reanalyses. These data sets are not only readily and easily available for future reanalyses, but are a valuable resource for research together with information based on the experience of the reanalysis efforts on their quality. In the next reanalyses additional efforts will be needed to gather and process correctly satellite radiance data and additional observation data sets not included in the first reanalyses. A special effort has been made to combine NCEP and ECMWF observation data sets; the first reanalyses revealed that each data set had observations not present in the other.

Volume of data use from reanalysis archives

It is useful to ask how much reanalysis data is being used. Data from the long NCEP/NCAR reanalysis project (data for 1948 — 99) are heavily used. During 1997 — 99, about 60% of the data being obtained from the Data Support (DSS) archives at NCAR were from this reanalysis project. The total use of data from DSS was 5600 GBytes in 1995 and 16,640 GBytes in 1999. The mode of delivery in 1999 was 2500 GBytes sent on tapes, 1600 GBytes sent on CD-ROMs, and 12,500 GBytes used on NCAR main computers by staff at NCAR and the universities. CD-ROMs have proved especially useful in countries outside North America. One CD-ROM was distributed to 13,000 meteorologists with the Kalnay *et al.* (1996) article in the Bulletin of the American Meteorological Society. Another CD-ROM will be distributed to more than 13,000 with the Kistler *et al.* (2000) article planned for the Bulletin and already has been distributed to 200 scientists at the 2nd International Conference on Reanalysis in Reading, England in August 1999.

Some of the data sent from NCAR goes into other archives around the world and is reissued from those archives. In addition, the NOAA Climate Diagnostics Center (CDC) in Boulder, Colorado, is a major distributor of reanalysis data via Internet. Other sites in the US and elsewhere are also acting as redistributors of NCEP Reanalysis data (e.g., the International Research Institute, IRI, at Lamont-Doherty, NY). NCEP archives provide data to NCEP personnel and also provide selected reanalysis and current CDAS fields to users outside NCEP via the Internet. In July 1997 for example NCEP's reanalysis web server received 21,000 hits. At that time, over 3000 computers had made hits on the web page. These computers included 135 US universities, 52 foreign countries, 35 US government organizations, 5 military organizations, and 282 .com or .net Internet domains. These distributions probably double the data distribution originating from NCAR.

4. A Strawman Proposal

In the following we provide a rough draft of a proposal that outlines the basic components, scientific requirements and development strategies, as well as the resources required for a national reanalysis program.

As outlined in the previous sections, we consider 10 years to be the appropriate time scale on which the U.S. should consider carrying out a major reanalysis production effort. It is important to realize, however, that the core of this program is not the production effort itself, but the necessary scientific analysis and technology development that will drive such production activities. These will be part of an on-going, iterative process, to improve our climate data sets.

We envision that the program will consist of three basic components. 1) A core Scientific Steering Group dedicated to work on high priority research and development focused on the Reanalysis. They would work with the major centers (e.g. NASA and NOAA) to define the data assimilation systems (DAS) to be used for reanalysis, provide basic DAS tools to the general community for testing, centralize the management of the reanalysis efforts, and ensure that the program has continuity and oversight. 2) Partnerships with selected organizations to carry out production, and to facilitate validation and data distribution, and 3) targeted community research and development efforts that specifically address the requirements of this program.

i. Scientific Requirements

The first generation of reanalysis efforts served to demonstrate the tremendous benefits of reanalysis to the Earth Sciences community (thousands of scientists throughout the world are using the reanalyses), and the lessons learned from these projects provide a valuable baseline from which to move forward. These efforts can be viewed as the first attempt by the climate community to capitalize on the successful development of data assimilation systems developed for operational numerical weather prediction (NWP). As such, these efforts have a strong heritage in NWP where the priorities have been on improving the accuracy of the analyzed atmospheric state. Refocusing these systems for Earth Science applications is just beginning. This requires new assimilation technologies that better address climate-specific issues such as model climate bias, observing system bias and inhomogeneities, handling of changes in the observing systems, physical consistency of budgets, improving estimates of physical processes, and developing methods for improving estimates of low frequency changes and trends, and uncertainties. The long-term challenge of reanalysis is to extend the analysis system to include all relevant components of the earth system.

In the following, we outline the near term priorities and indicate some of the long-term challenges of a reanalysis program. It is important to realize that the development of data assimilation for general Earth Science applications is in its infancy. As such, we cannot hope to produce the best possible climate product in one giant step. We must instead take incremental steps to address the major flaws of the current systems, and then get feedback from the broad spectrum of users.   At each step, we must weight the costs and benefits of carrying out a new reanalysis effort.

Priorities of a reanalysis program include:

1) Improved hydrological cycle, especially precipitation and clouds
2) better handling of changes in the observing systems, to minimize or eliminate spurious climate jumps in the reanalysis
3) improved estimates of uncertainty and bias, especially for the early, pre-satellite era, and for diagnostic fields
4) improved estimates of decadal and longer term changes.
5) increased resolution, critical for many applications.

Deficiencies in precipitation and clouds have limited the usefulness of the first generation of reanalysis products for a number of applications. For example, precipitation bias over summer continents (especially the diurnal cycle) impacts the usefulness of the data for driving land surface models, and other land hydrology studies.   Deficiencies in surface fluxes (largely due to cloud deficiencies) make current products poorly suited for many ocean-modeling applications. This also impacts their usefulness for general studies of the global water and energy budgets. These are critically important components for understanding global change and its impacts, and new reanalysis products must improve on these parameters. This should also lead towards the important goal of *coupled ocean-land-atmosphere reanalyses*.

Increased resolution is important for a number of applications. Key issues here involve improved resolution of land surface features, topography, the boundary layer, and the desire to resolve spatially confined phenomena such the Great Plains low level jet. The general problem of regional impacts of global-scale climate variability, such as the coupling between ENSO and regional monsoon features, will require analysis products with a spatial resolution that is at least 1 degree in latitude/longitude. The need to improve the resolution of land surface interactions, fronts, and hurricanes may eventually drive that requirement to much higher resolution (say $1/4^o$). However, these requirements must be weighted against the resolution of the input data. For example, during the early part of the data record the effective resolution of the observations may only be say 5^o so it may be advantages to run an *ensemble of assimilations* at lower resolution to help quantify *uncertainties in the reanalysis*.

One of the biggest long-term challenges is to improve estimates of decadal and longer-term variability and trends. Current reanalysis products provide much information on climate variability on interannual and shorter time scales, some useful information on decadal scale variability, and very little on longer term climate variations or trends. This largely reflects the time scale of observing system changes. It is currently very difficult to determine if any trends seen in the reanalysis data are real or whether they reflect the long-term changes in the observations (e.g. due to the introduction and improvements in satellite data, changes in number and quality of radiosonde observations). The reanalysis program should come up with a long-term strategy for minimizing the impact of observing system changes. That strategy should include observing system simulation studies, subsampling, efforts to reducing model bias, and carrying out multiple reanalyses with different systems. Estimating the slow changes (and in many cases small amplitudes) of the climate system on decadal and longer time scales may also require new approaches to data assimilation that, for example, take a much broader time window during each analysis step. Such a non-local analysis in time is analogous to current schemes that allow all observations in space to influence the analysis at a single point. Improved estimates of decadal and longer variability will allow us to address such issues as what generates multiyear to multidecadal variability in the tropical/subtropical climate and the North Atlantic Oscillation: can the atmosphere generate such variability intrinsically or is it generated by two-way, coupled ocean-atmosphere interactions, or by the oceans integrating atmospheric noise?

ii. Methodology

Reanalysis users realize that current reanalyses have many weaknesses that make it impossible to satisfy their many scientific requirements. These weaknesses represent limitations in the current technology and lack of basic research specifically required for the reanalysis. In order for the next generation reanalysis to be more useful than the first generation reanalysis, we must plan ahead and make every effort to solve a multitude of problems existing in the current systems. We also need to question whether it is a correct approach to utilize state-of-the-art operational analysis system for historical reanalyses, since the major purpose of operational centers is to make the best forecast and producing analyses climate is not their primary concern. Specifically, the focus of operational centers is the use of the latest high-resolution, high-accuracy, space-based observing systems: this may not be the best approach for the analysis of historical data.

The next generation reanalysis should be divided into at least two separate but closely connected analyses. This separation comes naturally from the different foci of users who utilize the reanalysis data. One reanalysis should aim at a long period going back to 1940's (hereafter called long reanalysis), and the other at focusing the post-satellite period (short reanalysis), where the observational coverage is more comprehensive, but has a different character. Both analyses should be continued to the present, since it has been clear that the NCAR/NCEP

Reanalysis is a more valuable community resource because of its real-time extension, CDAS[1].

Long reanalysis (1940s or earlier to the present)

The major objective of the long reanalysis is study of the long-term trend and decadal variations. It is essential that the analyses are not affected by the change in observational system, thus some sort of analysis based on a fixed observation system may be a natural choice. Sparse observational coverage in the pre-satellite period precludes the accurate analysis of small scales, so that only an accurate analysis of planetary and synoptic scales needs to be targeted. This focus allows us to use a relatively low-resolution analysis system on the order of T62 (200km).

The use of a limited number of observation to produce accurate planetary and synoptic scale analysis will be a new challenge. One logical approach to this difficult problem is the use of 4-dimensional variational data assimilation (4D-VAR) or a simplified Kalman Filter that can propagate information in space as well as in time. The development of such a system not only helps create accurate analysis in the pre-satellite period but also can promote a new tool in research and even contribute to the improvement of the operational analysis. It should be noted that such a system is already in use at ECMWF for their medium range prediction, thus the technique is known to work. The use of low-resolution analysis system will allow the use of fairly expensive 4-DVAR in the historical long analysis.

In parallel to this logical but rather laborious approach, it is necessary to pursue more conventional methods that may be possible using a currently available reanalysis system. The modification of the background error covariance matrix in the earlier analysis period has the potential for a significant improvement. Techniques for reducing the systematic biases in the forecast model should be developed. This would remove/help reduce trends induced by changes in the observation density. Observation system simulation experiments aiming at long-term trends and decadal phenomena, will be important for designing the fixed observation network. The portable reanalysis system developed at NCEP, the availability of short-and medium-range forecasts, and the quality-controlled data from the reanalysis are now available, so thathis research is already possible at various institutions and universities.

---

[1] It has been speculated that surface observations might be used to create useful reanalyses that went back to the late 1800's. Currently more research needs to be done on this subject, since if possible, this would greatly increase the usefulness of the reanalyses.

Since long reanalysis depends more critically on the small number of observational data, more stringent quality control of the observations is indispensable. The data need to have small bias and their characteristics need to be well documented during the analysis period. One of the outputs of the reanalysis is the observation meta-data that describe quantitative information on the quality of the observations. This product is under-utilized by researchers due to technical difficulties and their volume, but detailed study is essential before the next generation reanalysis begins.

Many reanalysis users were interested in the near surface conditions (precipitation, temperature, wind, and humidity). This is a weakness in the current systems as many of the surface observations are omitted or under utilized. Since many of the observations in the pre-satellite period are surface observations, research will be needed to implement this valuable data source into the assimilation system.

Short Reanalysis (1979-present)

The goal of the short reanalysis is to produce the best spatial analyses rather than the best analyses of the interannual variability. Abundance of observations, particularly the satellite and aircraft data, make it possible to aim at more accurate analyses of smaller spatial and temporal scales. The analysis should also be useful for short to medium range numerical weather prediction and longer-range seasonal predictions. The analysis must be accurate enough to be used in a variety of budget studies. Another application is various case studies that might require accurate analysis of small-scale details.

The analysis should utilize all the available observations in the framework of variational analysis, including the direct use of satellite radiance from various satellite instruments. Methods to use unconventional data, such as precipitation, OLR, planetary albedo, cloudiness, liquid water, aerosol and vegetation need to be further developed and refined. The improvement of the atmospheric model to reduce systematic error is also an important element for improving the initial guess as well as providing more accurate budget calculations. Coupling with comprehensive land and snow models needs to be pursued to improve analysis of near surface fields as well as land surface conditions. Development of variational analysis techniques for land conditions (soil wetness and snow) will provide more accurate surface conditions, which are essential for climate study. These requirements have many things in common with the objective of data assimilation research in progress at operational centers, and use of the state-of-the-art assimilation is justified for this particular analysis.

The short reanalysis places additional requirements on the time continuity of low frequency analysis. ECMWF ERA-15 demonstrated that unexpected change in analyzed climate may occur due to a change in the bias of satellite radiance observations. The new satellite observing systems have the potential of introducing unnatural changes. An example of a difficult problem that has could become even more important is the balance of mass and wind fields in the tropics. The stringent quality control of various satellite observations will be as important as the quality control of radiosonde observation in the long reanalysis, and because of the nature of observation and their amount, the task will be much more difficult.

One of the important elements missing in the first reanalyses was the estimate of the accuracy of the analysis. The uncertainty of the analysis can be accessed if multiple analyses are available. The use of multiple forecast models in the assimilation system is one practical method to generate such multiple analyses.

Finally, data collection, particularly precipitation, snow, soil wetness, ground temperature, cloudiness from surface stations need further efforts. Although ECMWF has started for their ERA-40 project, collection of various historical satellite raw radiances is an enormous task and continuing collection and quality control project is necessary.

Treatment of the stratosphere

Both the NCEP/NCAR and ERA-15 reanalyses included the lower-middle stratosphere, but had a rather low vertical resolution in this region, with only three to four model levels between 100 and 10 hPa. Analyses at 10 hPa were also impacted by the proximity of the upper boundary of the models. In the next national reanalysis, the stratosphere could be incorporated in different ways in the longer and shorter reanalyses:

(1) For the long reanalysis, the vertical resolution should be set to around 0.5-1.0 km (or higher) in the region between 300 and 50 hPa and at least 1 km between 50 and 10 hPa; the upper boundary of the model could be near the stratopause, with a few levels above 10hPa where there are few data. This would ensure that the tropopause region is adequately modeled for studies of the thermal structure and dynamical processes, as well as water vapor transport into the stratosphere.

(2) An additional 'stratospheric configuration' could be run for the short reanalysis period, when TOVS/SSU radiances and TOMS/SBUV ozone data are available. This should include a detailed troposphere and maintain a vertical resolution of 1-2 km up to the stratopause, with a low-resolution 'sponge layer' between 0.4 hPa (the uppermost level of SSU data) and 0.01 hPa (or less); ozone should be included as an assimilated variable, because of its impact on the total radiation budget and on local heating rates. This latter dataset would be extremely valuable for studies of long-term changes in ozone and off-line studies of trace gases in the middle atmosphere.

Coupling with ocean

The historical analysis of ocean is as important as the atmospheric analysis for climate study. For the next stage, having accurate surface winds and thermal forcings are important so that an ocean data assimilation can be driven by the atmospheric analyses. Producing more accurate surface winds is, dynamics might help as well a better use of surface observations.

Ultimately we should aim towards two-way *coupled ocean-land-atmosphere* data assimilation systems. One-way coupling, as presently carried out, still requires considerable research until full coupling becomes possible.

Special purpose reanalyses

The reanalysis project can produce special purpose reanalyses. The regional reanalysis over the US currently in progress at NCEP is an example one of such project. Land data assimilation (LDAS) is another possible reanalysis project if it were extended to a global domain an executed on the historical record. A stratospheric reanalysis has already been planned at NASA. Other possible reanalysis are regional analysis over meteorologically important areas (India, equatorial and western Pacific, Polar regions), coastal ocean and ocean wave analysis. Some of these can be carried out through international collaborations. Analysis of sea-ice, snow and long term evolution of vegetation may be also included.

III. Production, Validation, and Data Distribution

The US global reanalysis program requires significant resources. Numerous governmental, academic, and private institutions could be tapped to provide the necessary data, personnel, and program infrastructure to produce and analyze the reanalysis products. Data production would most likely be carried out in partnership with one or more of the major U.S. data assimilation centers (e.g. NOAA, NASA, Navy). There are also a number of candidate institutions with the capability to provide support for data distribution.

From the outset the relationship between those producing the reanalysis products and key product evaluation teams (e.g. for surface fluxes, precipitation), should be a partnership of the highest priority, founded on a principle of adequate support for activities constituting this core of the reanalysis program. The value of such a partnership will be evident through enhanced and quality-reviewed reanalysis products, efficient processing and demonstrable applications. Partnerships should also be encouraged between the resource providers and the larger community reanalysis product producers in order to foster widespread application and utility of the program products and knowledge. Several groups (e.g. NCAR Data Support Section, NOAA-NCEP, NOAA-CDC (Boulder, CO), NCDC) were instrumental in making the first NCAR/NCEP/NCDC reanalysis project a success. Their experience would be very valuable in developing a reanalysis program.

The integrity of the retrospective analysis of atmospheric data depends critically on the specified lower boundary conditions, in particular sea surface temperature (SST) and sea ice. Significant differences remain between current versions of global SST data sets, both in their long-term climatologies and in the monthly anomalies, especially before the advent of satellite data. Moreover, reanalysis should be eventually done with a coupled ocean atmosphere system, and linkages to efforts within the SST reanalysis teams and with ocean data assimilation groups at NOAA, NASA and the US Navy, as well as international efforts, are critical.

Validation strategies need to be developed the specifically address the quality requirements for different applications. In general, developers must work closely with the appropriate communities to help define the requirements in terms of error tolerances for specific applications. For example, for ocean modeling applications, knowing that a 10 W/m^2 error in the net surface heat flux can result in a 1C SST error helps to quantify error tolerances on assimilation cloudiness parameters and radiative fluxes. This illustrates the need for a continuing iterative process of reanalysis, examination, correction, new reanalysis, reexamination, correction, etc. that will become a natural component of a reanalysis program.

iv. Resources

At this stage, it is premature to try to provide more than very gross estimates of the anticipated costs for a national reanalysis program, which include both computer, data and human resources. We anticipate an approximately $8-12M/year program, with resources allocated to the three main components of the program as follow:

Core program:	40%
Partnerships:	20%
Community:	40%