The Importance of COADS for Global Reanalysis

Roy L. Jenne
National Center for Atrnospheric Research

Abstract

The existing daily atmospheric analyses for the period 1950-present are not of sufficient quality to permit the type of climate studies we need. There are several reasons for this: [1] there have been many changes of analysis procedures, [2] much of the observed data did not reach the operational centers in time for inclusion in the analyzed fields, and [3] in some cases the methods used for the analyses were not adequate. For example, it is only since about 1987 that the ocean surface-wind analyses have been good enough to drive ocean models, after some adjustment.

The COADS dataset that describes the surface marine conditions is of fundamental importance for reanalysis. It includes the world's collection of ship observations, drifting buoys, moored buoys, ice buoys, and the surface level from ocean station observations, e.g., XBTs. The COADS dataset has much more data than can be obtained in real-time.

The archives from reanalyses will contain data each six hours. There will be the normal fields such as temperature, humidity and winds. Boundary layer analyses will be archived. Many diagnostic fields such as precipitation, evaporation and surface radiation fluxes will also be available.

Introduction

As early as 1985 there were discussions about the possibility of making a reanalysis that would help achieve the goals of the TOGA Experiment. J. Shukla (University of Maryland) and others have helped to push this task forward. Some efforts arguing for a reanalysis project can be traced to as early as about 1982. From 1987-1991 there has been an increasing interest in making daily reanalyses of the global atmosphere. The data assimilation methods used to analyze the state of the atmosphere have shown major advances during the 1985-1991 period. A forecast is a part of these methods; therefore, the analyses improve when the forecast model is improved and when the methods improve. There have been large advances in the capability of forecast models and analysis methods. These improvements mean that the benefits of making a reanalysis have become much greater than before.

Bengtsson and Shukla (1988) published a paper that helped to start the movement toward planning for reanalyses. In early 1989 a small workshop was held to consider the initiative of making Tropical Ocean, Global Atmosphere (TOGA) reanalyses (Kinter and Shukla, 1989). There was a clear interest in analyzing the period 1979-1990 starting with two years in the 1980s (1982-1983) as a pilot project. Interest in pursuing this effort has grown, and now there is a NMC/NCAR project to reanalyze the entire period from 1958 to the present. An associated reanalysis workshop was held in April 1991 (Kalnay and Jenne, 1991). Representatives from ECMWF and the University of Maryland were present at the workshop; these two organizations are also planning to conduct their own reanalysis. As noted, when reanalyses efforts were first discussed in 1989, the main thrust was to redo the analyses for 1979 to 1989. Additional planning work has shown that it is technically possible, though not easy, to provide data inputs back to 1958, and later to 1950 or perhaps earlier. However, some networks of observations, such as Antarctica, did not start until 1956. Therefore, as we go back farther in time, data will not be available to help the analysis in some regions.

The data inputs for reanalyses include surface marine data, land surface observations, world rawinsondes, aircraft reports, satellite cloud drift winds and satellite sounders.

Is a Reanalysis Practical?

In order for a reanalysis to be practical, it first must be possible to prepare a sufficient quantity of input data. It also must be possible to sufficiently automate the monitoring of the analysis output so that 10 to 30 days of analysis can be accomplished during any one actual day. Finally, there must be enough computing time to permit a useful resolution for the analyses. A study was made to examine these questions (Jenne, 1988). It seemed technically possible to assemble the datasets; the experience of the FGGE analyses by GFDL showed that it was practical to accomplish about 30 days of analyses in one day.

Project to Prepare Data for Reanalysis

A project was needed to work on the necessary data inputs for global reanalysis. NCAR plans to use the many NMC archives (that are at NCAR) together with national archives of rawinsondes from several countries. The COADS marine dataset is also needed for the reanalyses.

NCAR has the key role to prepare data for reanalysis. We are being funded by NOAA and NSF for this work. NMC, ECMWF, and the University of Maryland will carry out the reanalyses. We have noted that NCAR has a joint proposal with NMC to prepare reanalyses for a 35-year period (1958-on). We are also working with the University of Maryland in support of their initiatives (to analyze 1982-1983). ECMWF plans to reanalyze the period 1979-present. The availability of data inputs for reanalysis and the history of the amount of effort needed to actually make the analyses is described in Jenne, 1988. More detailed plans for data preparation can be found in Jenne, 1991.

The initial two-year projects include preparation of global surface marine (COADS) data from 1947; world rawinsonde data from 1958; global surface land data from 1967; and world aircraft data from 1960.

Preparation of the Surface Marine Dataset (COADS)

All studies of long climate trends and variability of the ocean/atmosphere interface require the use of surface marine data. COADS is the world's best combined dataset of this type of data. From 1983 to 1984, NCAR, the Environmental Research Laboratories (ERL), and the National Climatic Data Center (NCDC) cooperated to prepare the COADS ship data (1854-1979). The data has been sent to 200 key research groups worldwide. Temporary, partial updates have been prepared by

ERL to cover the years after 1979. There is a critical need to fully update this dataset and a project to accomplish this is moving along.

MEDS (Marine Environmental Data Service of Canada) provided a lot of help to prepare and quality control the ocean buoy data used in the COADS update. The U.K. Meteorological Office is also now active in these efforts to prepare an improved ship dataset that will be merged with other COADS data sources.

Under the US-USSR data exchange, between Feb. 1989 and Sep. 1991 we have received a total of 25.2 million ship observations (see Yudin et al., 1992). These observations are mostly from Soviet ships. They cover a century-long period, but most data are from the period 1950 to 1990. These reports should provide a major addition to the existing COADS dataset which has 72 million reports for 1854-1979.

The COADS project involves merging about ten very large component datasets (see Woodruff et al., 1992). The components have to receive a series of corrections, be sorted, have duplicates eliminated, receive quality-control treatment, be resorted, and so forth. Statistics will then be calculated and a new round of products created. For use in reanalysis projects, the data will be sorted into a synoptic order (all global data for one day are together).

Status of this Project in June 1992

The task of converting many subsets of ship and buoy data to a common, format is proceeding. To speed the checkout task, two people independently convert each format and compare results. Portions of data for 1985-1989 have been sent to NMC to check out programs for analysis of sea surface temperature and for reanalysis.

Why have a COADS project? Why not just use data that can be collected in real-time? Table 1 shows that by waiting a few years the supply of ship data can be doubled. These counts of observations include only ships (not buoys, etc.). The reports on GTS are received in real-time. The data from logbooks arrive at a later time from many countries. Some of the delayed reports are the same as GTS reports; though, in general, logbook data are of better quality than GTS. The table below gives the counts of ship reports for a given year, after duplicates have been eliminated. Note that by waiting 5 years to receive delayed data, we were able to increase the number of reports for 1985 by 85%:

Table 1. Receipt of contemporary ship data, about 1985.

Approximate Unique Ship Reports	Percentage of Increase (relative to real time)	Delay Time
1,2000,000/year	--	Received in real time
1,856,000/year	55%	After 1 year
2,100,000/year	75%	After 2 years
2,220,000/year	85%	After 5 years

Notes:

(1) For a particular year (e.g., 1983), this shows the number of ship reports available after different periods of delay.

(2) Many reports are received as much as 10-30 years late.

Rawinsonde and Upper Wind Data

For Upper Air Data, NCAR will concentrate on providing rawinsonde data for 1958-1991. This will include data from the NMC decode for 1962 to the present. We also will provide world radiosonde observations (raobs) (May 1958 to April 1963) from the MIT general circulation project and place these data into one format. Data from some national archives is also available: US, Canada, Australia, Singapore, etc. We will provide a separate set of rawinsonde data for remote locations. In addition, about 10% of the effort will be devoted to getting data earlier than 1958. After 1992, an increasing amount of effort will be spent on the early data. We will also provide a better set of aircraft data from at least 1960 on.

Surface Land Data

For land synoptic data, we will concentrate on providing data for 1967 to 1992. Some effort will be put into providing the earlier data for 1948-1967, but this data probably will not be ready for the first reanalysis. The land synoptic data are not as critical as ocean surface observations, because there is often a reasonable density of rawinsonde observations over land.

Philosophy of Reanalysis

The reanalysis should not be undertaken as a one-time project that is good for all time. We know that assimilation methods will gradually improve with time, and that over the next 5 to 12 years, we can improve the quantity and quality of input data. The plan calls for incremental data improvements with time.

Under the proposal, NMC/NCAR will analyze the atmosphere for test periods to ensure that the whole system is working and then analyze long periods (1958-1990) using an atmospheric model. To analyze the ocean, the ocean model will use forcing fields from the atmospheric analyses, COADS data for the surface ocean, and ocean subsurface data for the deeper ocean.

We will probably make an independent updated analysis of sea surface temperature for the whole reanalysis period. Another analysis of the full period of data will be made about three to five years later (with improved model and data inputs). By then, observed data will be available for earlier years.

Strategy for Preparing Data Inputs for Reanalysis

Our goal at NCAR is to provide data inputs significantly better than those used in real time, and to have the data ready in a timely fashion. Other improvements can be made available at later dates. We have defined major subgoals for each type of data. Goals were determined for Reanalysis data (Version 1 to be ready about June 1993) and Version 2 (to be ready in 1996). The initial data set will not be as complete as we would like.

We will need to merge many datasets to prepare data of major types (surface marine, surface land, rawinsonde and winds aloft, aircraft data and so forth). However, we will not merge the major types. This strategy will avoid problems such as those experienced with FGGE (First Global Atmospheric Research Program [GARP] Global Experiment), where it became hard to correct errors or access subsets.

Scope of the Work

It is a very large task to prepare the major types of data. However, automation helps control the amount of work needed. In 1982, we estimated that 5,000 computer batch jobs at NCAR would be needed to accomplish the COADS project. With this number of tasks, we knew that the job could still not be completed on time and would experience human errors, unless we used automation. Therefore, we invented a set of procedures that we call “run control automation” that has streamlined the process.

In the overall context, this project entails preparing datasets with nearly all of the world’s synoptic scale observations taken over the past 35 to 45 years. We will capitalize on the results of other large tasks such as COADS and will use the results of data gathering done in the last 25 years. We will use the efforts of national data centers in various countries (we already have a number of these datasets). ERL, NCDC, Canada, and others will contribute to this project.

Advantage of a Reanalysis Compared with Summary Statistics

A good data assimilation scheme (such as the methods to be used for reanalysis) can extract more information from the observations than is possible by summarizing data on 1° or 2° squares. There are several reasons for this. A ship pressure (or wind), for example, is correlated over long enough distances that on a given day; it can contribute information to nearby grid boxes, yet in statistical summaries the data is only used in one box. The reanalysis methods properly use the information in adjoining boxes. The assimilation methods include a good forecast model. Even if there is no data-in part of the northern Pacific, the analysis will probably be good, because the forecast model can make a good prediction based on data from Asia, Alaska, and from satellites.

We might ask whether there is still a value to the type of statistical summaries used in COADS. The answer is "yes." When data are very sparse, the daily analyses may not have enough information to adequately describe the synoptic scale systems. In this case, the output from the model can be biased toward the natural model statistics, rather than the statistics of the atmosphere. The COADS type of statistics can be used to help verify the output of the analyses. In the last few decades, there should be enough global observations (surface and upper air) that the output from the reanalysis will be very good. But in early years without satellite help, part of the S. Hemisphere may be difficult to analyze well.

Use of CD-ROM and Tape Technologies to Distribute Data

The first CD-ROMs with data for the geosciences were produced in mid-1987. Now, a flood of them is being prepared. A CD-ROM holds about 650 MB, compared to 125 on a high-density half- inch tape. The access time to any part of the disk is about 0.4 seconds. A reader costs about $400 to $600. The 4 mm digital audio tape drives (DAT) can now be purchased for about $1,600. We hope that the price will come down to $900-$1,200 in one to two years. These small tapes hold 1300 MB of data. Exabyte tape drives hold more data than a DAT tape, have a faster data rate and cost as little as $2,400-$3,500. These technologies will help with the distribution of COADS data, and data from the reanalyses.

References

Bengtsson, L. and J. Shukla, 1988: Integration of Space and In-situ Observations to Study Global Climate Change, *Bulletin of the American Meteorological Society*, **69**, 1130-1143.

Jenne, R., 1988; A Global Reanalysis: Data Inputs and Methods, *Proceedings of the 13th Annual Climate Diagnostics Workshop*, NOAA, CAC, pp. 79-84.

Jenne, R., 1991: Initiative to Prepare Data Inputs for Reanalyses, Internal NCAR Report, 59 pp.

Kalnay, E. and R. Jenne, 1991: Summary of the NMC/NCAR Reanalysis Workshop of April 1991, *Bulletin of the American Meteorological Society,* **72**, 1897-1904.

Kinter, J. and J. Shukla, 1988: Reanalysis for TOGA, Report of Panel meeting February 1989, University of Maryland, College Park, MD. This document is summarized in the *Bulletin of the American Meteorological Society*, November 1989.

Woodruff, S.D., S.J. Lubker, and M.Y. Liu, 1992: Updating COADS—Problems and Opportunities. *In*: Diaz, H.F., Wolter, K, and Woodruff, S.D. (eds.), *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1993*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume)

Yudin, K.B., I.G. Ylyanich, V.N. Popova, and Ye. M. Krakanovskaya, 1992: Russian Marine Meteorological Data Set. *In*: Diaz, H.F., Wolter, K, and Woodruff, S.D. (eds.), *Proceedings of the International COADS Workshop Boulder, Colorado, 13-15 January 1993*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume)