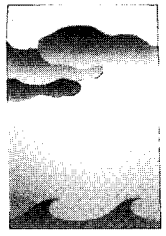
The NCEP/NCAR 40-Year Reanalysis Project



E. Kalnay,* M. Kanamitsu,* R. Kistler,* W. Collins,* D. Deaven,* L. Gandin,* M. Iredell,* S. Saha,* G. White,* J. Woollen,* Y. Zhu,* M. Chelliah,+ W. Ebisuzaki,+ W. Higgins,+ J. Janowiak,+ K. C. Mo,+ C. Ropelewski,+ J. Wang,+ A. Leetmaa,* R. Reynolds,* Roy Jenne,# and Dennis Joseph#

ABSTRACT

The NCEP and NCAR are cooperating in a project (denoted "reanalysis") to produce a 40-year record of global analyses of atmospheric fields in support of the needs of the research and climate monitoring communities. This effort involves the recovery of land surface, ship, rawinsonde, pibal, aircraft, satellite, and other data; quality controlling and assimilating these data with a data assimilation system that is kept unchanged over the reanalysis period 1957–96. This eliminates perceived climate jumps associated with changes in the data assimilation system.

The NCEP/NCAR 40-yr reanalysis uses a frozen state-of-the-art global data assimilation system and a database as complete as possible. The data assimilation and the model used are identical to the global system implemented operationally at the NCEP on 11 January 1995, except that the horizontal resolution is T62 (about 210 km). The database has been enhanced with many sources of observations not available in real time for operations, provided by different countries and organizations. The system has been designed with advanced quality control and monitoring components, and can produce 1 mon of reanalysis per day on a Cray YMP/8 supercomputer. Different types of output archives are being created to satisfy different user needs, including a "quick look" CD-ROM (one per year) with six tropospheric and stratospheric fields available twice daily, as well as surface, top-of-the-atmosphere, and isentropic fields. Reanalysis information and selected output is also available on-line via the Internet (http//:nic.fb4.noaa.gov:8000). A special CD-ROM, containing 13 years of selected observed, daily, monthly, and climatological data from the NCEP/NCAR Reanalysis, is included with this issue. Output variables are classified into four classes, depending on the degree to which they are influenced by the observations and/or the model. For example, "C" variables (such as precipitation and surface fluxes) are completely determined by the model during the data assimilation and should be used with caution. Nevertheless, a comparison of these variables with observations and with several climatologies shows that they generally contain considerable useful information. Eight-day forecasts, produced every 5 days, should be useful for predictability studies and for monitoring the quality of the observing systems.

The 40 years of reanalysis (1957–96) should be completed in early 1997. A continuation into the future through an identical Climate Data Assimilation System will allow researchers to reliably compare recent anomalies with those in earlier decades. Since changes in the observing systems will inevitably produce perceived changes in the climate, parallel reanalyses (at least 1 year long) will be generated for the periods immediately after the introduction of new observing systems, such as new types of satellite data.

NCEP plans currently call for an updated reanalysis using a state-of-the-art system every five years or so. The successive reanalyses will be greatly facilitated by the generation of the comprehensive database in the present reanalysis.

*Environmental Modeling Center, National Centers for Environmental Prediction, Washington, D.C.
+Climate Prediction Center, National Centers for Environmental Prediction, Washington, D.C.
#National Center for Atmospheric Research, Boulder, Colorado
Corresponding author address: Dr. Eugenia Kalnay, director, Environmental Modeling Center, W/NMCZ, Room 204 WWB, Washington, DC 20233.
E-mail: WD23EK@SUN1.WWB.NOAA.GOV
In final form 5 September 1995.

1. Introduction

The National Centers for Environmental Prediction [NCEP, formerly known as the National Meteorological Center (NMC)]/National Center for Atmospheric Research (NCAR)[1] Reanalysis Project began in 1991 as an outgrowth of the NMC Climate Data Assimila-

[1]A list of acronyms used in this paper is included in appendix D.

tion System (CDAS) project. The motivation for the CDAS project was the apparent "climate changes" that resulted from many changes introduced in the NMC operational Global Data Assimilation System (GDAS) over the last decade in order to improve the forecasts. These jumps in the perceived climate parameters obscure, to some extent, the signal of true short-term climate changes or interannual climate variability. An obvious example is presented in Fig. 1, which shows large jumps in the analyzed virtual temperature at 1000 hPa in the Pacific Ocean when the model was changed. The impact of system changes on other parameters, such as estimated precipitation and its distribution, is more subtle and therefore harder to separate from the true climate anomaly signals.

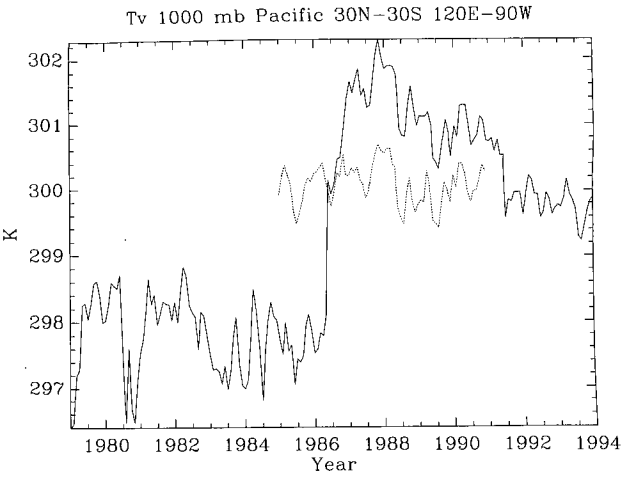


FIG. 1. Trace of the 1000-hPa virtual temperature averaged for the tropical Pacific ocean in the NCEP operational Global Data Assimilation System (solid line), showing the impact of changes in the model and in the reanalysis (dotted line).

The basic idea of the reanalysis project is to use a frozen state-of-the-art analysis/forecast system and perform data assimilation using past data, from 1957 to the present (reanalysis). Moreover, the same frozen analysis/forecast system will be used to continue to perform data assimilation into the future (CDAS) so that climate researchers can assess whether current climate anomalies are significant when compared to a long reanalysis without changes in the data assimilation system. In addition, there will be a one-way coupled ocean reanalysis, in which the surface fluxes from the atmospheric model will be used for the ocean data assimilation. The NCEP/NCAR 40-year reanalysis should be a research quality dataset suitable for many uses, including weather and short-term climate research.

The project development has been supported by the National Oceanic and Atmospheric Administration's (NOAA) Office of Global Programs. An advisory panel chaired by J. Nogués-Paegle guided it throughout the developmental period (1989–93). After the execution phase started in 1994, the advisory panel was replaced by a users' advisory committee, chaired by A. Oort. The reanalysis system was designed at NCEP, with the participation of over 25 scientists from NCEP's Environmental Modeling Center, Climate Prediction Center (CPC), the Coupled Model Project, and Central Operations. Scientists at NCAR performed most of the data collection and obtained many special datasets from international sources that were not available operationally through the Global Telecommunications System (GTS). E. Kung (University of Missouri) acquired early data from China. We also had the collaboration of NOAA/National Environmental Satellite, Data and Information System (NESDIS), who provided the TIROS (Television Infrared Observation Satellite) Operational Vertical Sounder (TOVS) data; the Geophysical Fluid Dynamics Laboratory (GFDL) for the ocean reanalysis; the United Kingdom Meteorological Office (UKMO), who will supply their global ice and SST reanalyses (GISST) for the earlier periods; the Japanese Meteorological Agency (JMA), who provided cloud-track winds and special rawinsonde data not available on GTS; and the European Centre for Medium-Range Weather Forecasts (ECMWF), who filled some data gaps and provided a sea ice database. The National Aeronautics and Space Administration/Goddard Laboratory for Atmospheres (NASA/GLA) has provided retrievals missing from the NCEP archives and offered to perform TOVS retrievals for several months of missing data in 1986. The NOAA/Environmental Research Laboratory (ERL)/Climate Diagnostic Center has provided funding for archival and tape handling development and support.

The early design of the project was discussed in an NMC/NCAR reanalysis workshop that took place at NMC in April 1991 (Kalnay and Jenne 1991). The workshop had the participation of representatives of all the groups planning to perform reanalyses at that time [Center for Ocean–Land–Atmosphere Interactions (COLA), ECMWF, and NASA/GLA], as well as of the major types of users (e.g., for short- and long-term dynamics and diagnostics, transport of trace gases, climate change, predictability, angular momentum and length of day, coupled models, etc.). The near-final design was reviewed in October 1993 by the advisory committee, who suggested several addi-

tional tests and modifications before the start of the operational phase (started in May 1994). Representatives of the major agencies interested in the project [NOAA, National Science Foundation (NSF), NASA, Department of Energy (DOE)] and of the other groups performing reanalyses also participated in the review of the NCEP/NCAR project. The other plans presented in the October 1993 review were those of ECMWF, which is performing a 15-yr reanalysis for 1979–94; NASA/GLA (Schubert et al. 1993), which performed a reanalysis for 1985–90, the U.S. Navy (1985 to the present); and COLA, which performed an 18-mon reanalysis for the 1982/83 El Niño. Such multiple reanalysis projects offer a great opportunity for cooperation and intercomparison, which should enhance each of the projects. In particular, the NCEP has benefitted from the COLA project through the transfer of the Gridded Analysis and Display System (GrADS), which has greatly enhanced the NCEP's developmental graphical display system, and from the close interaction with NASA and ECMWF scientists performing a similar reanalysis.

The purpose of this paper is to update the documentation of the NCEP/NCAR system design, output, and plans for distribution. The basic characteristics of the system are summarized in section 2, and the data to be used in section 3. The three modules of the reanalysis system (data quality control preprocessor, analysis module with automatic monitoring system, and output module) are described in sections 4, 5, and 6, respectively. The CDAS, which uses the same frozen system but continues the analysis into the future, is discussed in section 7. Section 8 summarizes the coupling with the ocean reanalysis. Section 9 contains an assessment of the reliability of the reanalysis output and the impact of changes in the observing system. Section 10 summarizes the project. More detailed documentation is available from the NCEP (Kalnay et al. 1993).

2. Overview of the NCEP/NCAR reanalysis system and execution plan

The NCEP/NCAR Reanalysis Project has two unique characteristics: the length of the period covered and the assembly of a very comprehensive observational database. The reanalysis will cover the 40-year period 1957–96 and will continue into the future with the CDAS. The observations will be saved in the World Meteorological Organization's (WMO) binary universal format representation (BUFR), with additional information, such as the first-guess and quality control decisions, incorporated into the report. We are also considering the feasibility of extending the reanalysis back to 1946, when the Northern Hemisphere (NH) upper-air network was established, as suggested by several researchers. The length of the reanalysis, and the desire to carry it out as quickly as possible to increase its usefulness, led us to design a system able to perform one month of reanalysis per day. Such a fast pace of execution required the development of a reanalysis system much more robust and automated than the analysis-forecast systems used for operational numerical weather prediction. As a result, the NCEP/NCAR reanalysis system has many novel features not yet present in operational or research numerical weather forecasting systems.

As shown in Fig. 2, the NCEP/NCAR reanalysis system has three major modules: data decoder and quality control (QC) preprocessor, data assimilation module with an automatic monitoring system, and archive module. The central module is the data assimilation, which has the following characteristics:

- T62 model (equivalent to a horizontal resolution of about 210 km) with 28 vertical levels. The model is identical to the NCEP global operational model implemented on 10 January 1995, except for the horizontal resolution, which is T126 (105 km) for the operational model (Kanamitsu 1989; Kanamitsu et al. 1991).
- Spectral statistical interpolation (SSI, or 3D variational) analysis, with no need for nonlinear normal-mode initialization (Parrish and Derber 1992; Derber et al. 1991); improved error statistics; and the balance constraint on the time derivative of the divergence equation implemented at NCEP in January 1995 are also included (Derber et al. 1994).
- Complex QC of rawinsonde data, including time interpolation checks, with confidence corrections of heights and temperatures (Collins and Gandin 1990, 1992); Optimal interpolation (OI)-based complex QC of all other data (Woollen 1991; Woollen et al. 1994).
- Optimal averaging of several parameters over a number of areas, providing more accurate averages and estimates of the error of the average (Gandin 1993).
- Optimal interpolation SST reanalysis (Reynolds and Smith 1994) starting from 1982 and UKMO GISST for earlier periods.

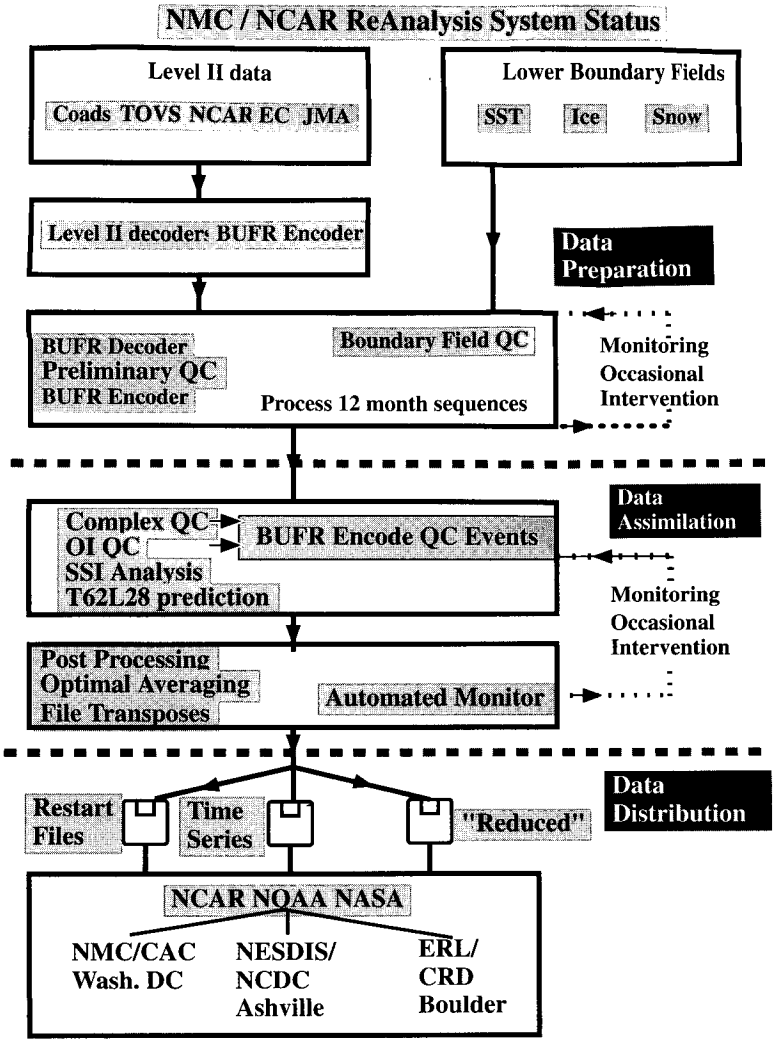


FIG. 2. Schematic of the main components of the NCEP/NCAR reanalysis system and their state of readiness on January 1995. "Underway" means that the component is working but it still being improved. ("NMC" has changed to "NCEP.")

- One-way coupled ocean model 4D assimilation for 1982 onward (Ji et al. 1994).
- The same CDAS will be used into the future, from January 1995 onward.

To support a rate of reanalysis of about one month per day, it is necessary to ensure that the data input is generally free of major data problems such as wrong dates, wrong locations, garbled information, etc., for both conventional and remotely sensed data. This is particularly important for old data, which have not been previously used at NCEP. Similarly, the rate of one month of reanalysis per day does not allow for the detailed human scrutiny that operational output normally receives. For this reason, we created a data quality control preprocessor and an analysis output QC monitoring module (Fig. 2): The data input is preprocessed, and all the analysis output fields are monitored with a "complex QC" monitoring system, in which the statistics of the data, time tendencies, etc. are compared to climatological statistics in order to detect errors. These statistics include tendency checks. (These monitoring systems will also be implemented operationally, and their development constitutes a major spin-off from this project for NCEP.)

It was decided early in the project that one type of output could not satisfy the needs of the many different types of users. For this reason the output module allows for several different archives:

- Level-2 observational data in BUFR, including QC, climatological, analysis, and 6-h forecast information.
- Comprehensive analysis, first-guess, and diagnostic fields presented in "synoptic" form (all fields every 6 h) in the model sigma coordinates, as well as in pressure and isentropic coordinates, in gridded binary (GRIB) format. A restart file is included once a month to allow for rerunning shorter periods with enhanced diagnostics.
- A time series archive in which each field is available for all times, including standard pressure level fields, precipitation, surface fluxes, and other widely used diagnostic fields. This format will be the most useful for many users.
- A "quick look" archive on CD-ROM, one per year, including the most widely used fields: daily values of variables at selected tropospheric and stratospheric pressure levels, surface and top-of-the-atmosphere fluxes, precipitation, monthly and zonal averages of most fields, covariances, isentropic level variables, etc. A special CD-ROM, containing 13 years of selected observed, daily, monthly, and climatological data from the NCEP/NCAR Reanalysis, is included with this issue (appendix E).

- Eight-day forecasts performed every 5 days, which should allow predictability studies and estimates of the impact of inhomogeneities in the observing systems coverage, with anomaly correlation scores.
- A subset of the output is posted on the NCEP public server and is available through anonymous FTP.
- NCAR, National Climate Data Center (NCDC), and CPC will distribute the bulk of the reanalysis data.
- Reanalysis information and selected output is also available on-line via the Internet (http// :nic.fb4.noaa.gov:8000).

An important question that has repeatedly arisen is how to handle the inevitable changes in the observing system, especially the availability of new satellite data, which will undoubtedly have an impact on the perceived climate of the reanalysis. Basically, the choices are (a) to select a subset of the observations that remains stable throughout the 40-yr period of the reanalysis or (b) to use all the available data at a given time. Choice (a) would lead to a reanalysis that has the most stable climate, and choice (b) to an analysis that is as accurate as possible throughout the 40 years. With the guidance of the advisory panel, we have chosen (b), that is, to make use of the most data available at any time. However, in order to assess the impact of the introduction of new observing systems on the perceived climate of the reanalysis, we have decided to produce a parallel reanalysis, at least 1 year long, without using a large new observing system. This will allow the users to assess the extent to which the new observing system influences the perceived climate and the annual cycle.

The execution phase started in May 1994 on the Cray YMP 8 supercomputer provided by the NCEP for this project. About 24 h of the CRAY YMP (2–7 processors) are needed in order to perform one month of reanalysis and forecasts per day. By September 1995, 13 years (1982–94) should be completed (in addition to several years of reruns performed to assess the impact of changes in observing systems, to correct problems discovered in the database, etc.). Next, the period 1979–82 will be reanalyzed and completed around the end of 1995, followed by the 1957–78 decades. We expect to complete the 40 years of reanalysis (1957–96) by early 1997. The extension into 1948–57, if feasible, would be done during 1997.

This first phase of reanalysis will be followed by a second phase in which a 1998 state-of-the-art system will be used for a second reanalysis. NCEP plans currently call for an updated reanalysis every 5 years or so. The successive reanalyses will be made easier by the availability of the comprehensive database in BUFR generated by the present reanalysis.

3. The preparation of data for reanalysis

The data collection is a major task that has been performed mostly at NCAR. Surface and upper-air observations are being prepared for the reanalysis. The plan is to use the data available for the original operational NCEP analyses (available from March 1962 on) and to add other datasets to capture the older data from about 1948 on. Additional data inputs for 1962 on will provide much more data than was first available operationally and will be merged and formatted in BUFR at NCEP. The component datasets are listed below. For further details consult Kalnay et al. (1993) and the extended texts provided by NCAR (see list of texts below).

a. Global rawinsonde data

NCAR has tapes of the NCEP GTS data with upper-air observations from March 1962 on, which will be the main data source for reanalysis. We plan to provide both the GTS data (which also has pibals and aircraft) and also raobs from national archives in various countries. NCAR has raobs received directly from some countries such as South Africa, Australia, Canada, Argentina, Brazil, the United Kingdom, France, and from the United States (NCDC). The U.S. Air Force (USAF) prepared a global collection of data (TD54) that is mostly for the period 1948–70, which will be included. GFDL is helping with processing and checking this set, which will all be ready for the first reanalysis. The University of Missouri (E. Kung) is collaborating with some of the checking between different sources of the same data and has the obtained daily upper-air data of 30 stations over China from the Chinese State Meteorological Administration for the period of 1954–62. Under the United States–Russia bilateral exchange effort led by R. Jenne (NCAR) and S. Shumbera (NCDC), the United States has received 20 magnetic tapes with upper-air data from 57 USSR stations for 1961–78. The Japanese Meteorological Agency has provided NCEP with additional data not available over GTS.

During the reanalysis, it was found that the count of significant level winds was low from August 1989 to September 1991 in the NCEP tapes but not over the United States and China. ECMWF supplied their data

to fill the gap. Interestingly, the ECMWF had a similar, but complementary, low count over the United States and China.

NCAR, NCDC, Russia, Europe, and other organizations, including WMO, have interests in improving the global archive of rawinsonde data. We anticipate various collaborations to improve the basic input sets and to accomplish merges. However, their results will be available for later reanalyses, not for the first one.

b. COADS surface marine data

The Comprehensive Ocean–Atmosphere Data Set (COADS),[2] first released in 1983 and recently updated, includes ships, fixed buoys, drifting buoys, pack-ice buoys, near-surface data from ocean station reports (XBTs, etc.), and other data. An update for 1980–93 has been completed, and work is progressing on all the surface marine data for 1947–79.

c. Aircraft data

Aircraft data is available from the NCEP GTS source starting in March 1962. Additional data have been gathered from several sources, including data from New Zealand for February 1984–June 1988, some of which did not get into GTS. Aircraft data from experiments such as the GARP (Global Atmospheric Research Program) Atlantic Tropical Experiment (GATE, summer 1974) and the First GARP Global Experiment (FGGE) (1979) will be used. Selected USAF reconnaissance data is available, starting from 1947. Data from Tropical Wind Energy Conversion Reference Level Equipment (TWERLE) constant-pressure balloons for the Southern Hemisphere (SH, July 1975–August 1976) will be in the dataset. These balloons provide data similar to a single-level rawinsonde near 150 mb.

d. Surface land synoptic data

Global GTS data (usually every 3 h) are available starting from 1967 from air force or NCEP sources. Earlier years are available from the air force tape deck 13 and from U.S. hourly data (from NCDC). The data coverage is fairly good from 1949 on.

e. Satellite sounder data

The basic radiances are available for the following periods:

- SIRS[a] IR sounders — Apr 1969–Apr 1971
- SIRS on early NMC tapes (not radiances) — Nov 1969–Sept 1992
- VTPR[a] IR sounders — Nov 1972–Feb 1979
- TOVS sounders (HIRS, MSU, SSU)[a] — Nov 1978–present
- HIRS data test system — Aug 1975–Mar 1976

In the first phase of the reanalysis we plan to use the original operational TOVS retrievals of NESDIS (2.5° space resolution). A system based on the 3D variational assimilation of variances will be used in the second phase of the reanalysis (to start in 1998). It should be noted that the pilot experiments comparing reanalysis with and without the use of satellite data, to be discussed in section 9, have provided useful information regarding the uncertainties of the analysis without satellite data. This is very important for the period before 1979 when no TOVS satellite soundings were available. We hope to assimilate VTPR and HIRS data available before 1979 for the Southern Hemisphere, although we have no recent experience with that data, and there may be unforeseen problems.

f. SSM/I surface wind speeds

Special Sensing Microwave/Imager (SSM/I) data became available in July 1987, and at NCEP the assimilation of surface winds became operational on 10 July 1993. We adopted the neural network algorithm of Krasnopolsky et al. (1995), which results in wind speeds significantly closer to buoys wind speeds and with better coverage under cloudy conditions than the present operational algorithm used at NCEP. We initially used a subset of the high-resolution SSM/I radiance data archived by NESDIS for climate purposes. However, after over 4 years were reanalyzed, several problems were found that indicated that it would be necessary to use the original dataset. The high volume of these data (much larger than all other data together) also resulted in a significant slowdown of the reanalysis. For this reason it was decided that the first phase of the reanalysis will not include wind speeds from SSM/I (except for limited data impact studies). We plan to use the SSM/I wind speed, as well as total precipitable water and other parameters, in the second phase of reanalysis.

g. Satellite cloud drift winds

Satellite cloud drift winds are used from the origi-

[2]COADS is a joint project of NOAA/ERL, NCAR, and NCDC. Many other organizations and countries have also contributed to its creation.

[a]See appendix D.

nal NMC tapes and from the Geostationary Meteorological Satellite (GMS) cloud drift winds received from the JMA for the period 1978–91.

A text entitled "Data for Reanalysis: Inventories"[3] has various maps and displays that illustrate the typical coverage of surface and upper-air data that are already available. Most of this information covers the period from about 1948 on. The coverage of data is rather encouraging, even for the earlier years. We note, however, that rawinsonde observing networks for Antarctica and the west coast of South America did not start until July 1957. Many other reports have been prepared that give more information about the attributes of different datasets and the status of projects to prepare the data. Papers have been prepared that focus on different issues; a selection of these papers is given below. Additional papers are available upon request.

Selection of texts about data for reanalysis (contact NCAR for further information):

Text	Date
Data for reanalysis: Inventories	Nov 1992
Sea surface temperature data	1 Feb 1993
Sea ice data	2 Apr 1993
Rawinsonde data for reanalysis	24 Oct 1994
Dataset of tropical storm locations	26 Jan 1993
NMC upper-air data: 1962–72	29 Mar 1994
Global satellite sounder data	12 Aug 1994
Surface land synoptic data	May 1994
Ice cap buoy update	5 Apr 1994
Inventories of data for reanalysis	Mar 1995
Analyses for the SH from 1951 on	18 Mar 1993
Status of reanalysis data	1 Apr 1993

4. Data preprocessor

The purpose of the preprocessing reanalysis module (Fig. 2) is to reformat the data coming from many different sources (Fig. 3) into a uniform BUFR format and to preprocess 1 or more years at a time, before the actual reanalysis module is executed at the rate of one month per day. This allows detection of major data problems with sufficient lead time (a few days before the execution of the reanalysis), so that human monitors can try to take corrective action. The preprocessor thus minimizes the need for reanalysis reruns due to the many data problems that frequently appear, such as data with wrong dates, satellite data with wrong longitudes, etc. The preprocessor also includes preparation of the surface boundary conditions (SST, sea ice, etc.).

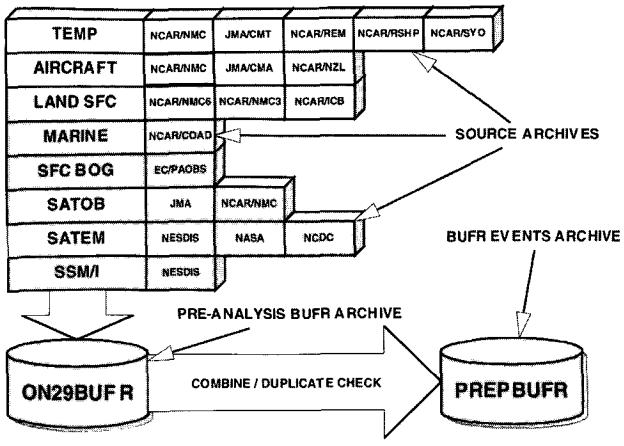


FIG. 3. Schematic of the data input archive structure for the reanalysis. The data are classified into eight basic types (rawinsondes, aircraft, land surface, marine, surface bogus, satellite temperature soundings, cloud-tracked winds, and SSM/I). The PREPBUFR archive adds "events" on each datum as it flows through the reanalysis.

a. Satellite data

A special satellite TOVS soundings data monitoring system has been developed. It is intended as a quality control of the data in the NESDIS archive tapes that can suffer from errors in dates and orbits not likely to occur in the daily operational products. Satellite data in grid boxes of 10° × 10°, as well as single satellite observations, are quality controlled. The average in the box, the variance in the box, and the absolute value of tendency of the box average are compared with a climatology to flag suspicious groups of satellite data.

b. CQC with temporal check

The complex quality control (CQC) system (described in the next section) is included in the prepro-

[3]Various documents describing this dataset are available from R. Jenne, NCAR, P.O. Box 3000, Boulder, Co 80307-3000.

cessor but without the use of the first guess of the model. The baseline check in the preprocessor (see next section) allows for the detection of changes in the station locations, an important problem that interferes with the accurate detection of climate change.

c. Climatological QC test of data

The automatic monitoring system developed for the reanalysis output (section 8) is based on climatological tests with three-dimensional (grid point) statistics computed for each month. The space-time character of the statistics proved to be very successful in finding problems in the pilot experiments for the reanalysis (section 8), which were then related to unusual data errors, leading also to corrections of several errors present in the operational system. This led us to check the data directly within the preprocessor, by expressing the observation anomaly in units of standard deviations with respect to climatology, a number which can be generated from the BUFR "events" archive (see section 5f). Such a check allows human monitors to check for unusual data present in unusual amounts, before the execution of the monthly reanalysis, and provides the optimal interpolation quality control (OIQC) with additional information that can be used by its decision making algorithm (DMA) as input to the reanalysis.

d. Boundary fields

The following analyses and climatologies are used for the boundary fields:

1) SST: Reynolds reanalysis for 1982 on, when AVHRR data became available, and the UKMO GISST for earlier periods (D. Parker 1992, personal communication).
2) Snow cover: NESDIS weekly analyses and climatology, updated weekly (D. Garrett 1995, personal communication);
3) Sea ice: The ice field derived from SMMR/SSMI data, and quality controlled by B. Nomura for the ECMWF reanalysis, has also been adopted at NCEP for the period 1979–93. Beyond 1993, a similar algorithm developed by R. Grumbine is being used. For earlier periods we plan to use the analysis from Joint Ice Center analyses when available, J. Walsh and GISST analyses otherwise. These have been incorporated into the SST analysis so that all values below −1.8°C are considered sea ice. R. Grumbine has inserted a more realistic glacial coverage for the Ross Ice Shelf and other regions of the Antarctic. A simple check (comparison with monthly climatologies and standard deviations) should help to ensure that no major errors are present in the data or made inadvertently during their use.
4) Albedo: Matthews (1985).
5) Soil wetness: Updated during the analysis cycle. The model uses the Pan and Mahrt (1987) and Mahrt and Pan (1984) soil model. There is no nudging of the soil moisture using concurrent data, and a very small coefficient (0.05) is used to nudge the soil moisture toward climatology. Soil moisture fields show interannual variability but no long-term drift (Fig. 7).
6) Roughness length: From SiB.
7) Vegetation resistance: From SiB (Dorman and Sellers 1989).

However, preliminary reanalyses showed that the original resistance over regions deemed to be covered by winter wheat had excessively high plant resistance in the summer and fall, resulting in temperatures that were too high and low precipitation in the eastern North America summer (S. Saha and H.-I. Pan 1994, personal communication). For this reason, we are using the minimum monthly resistance value for each grid point. Monthly climatologies are the backup of the analyzed fields when these updated fields are not available.

5. Data assimilation module

a. System configuration

The CDAS/reanalysis is executed at the NOAA Central Computer Facility in Suitland, Maryland. Unlike the operational NCEP system, which currently is based on both IBM–MVS-type and Cray–UNIX computers, in the CDAS/reanalysis system all processing is done in the Cray-UNIX environment. Observations will be encoded in BUFR, and gridded data in GRIB the standard WMO formats. This system will soon be also adopted by NCEP for its normal operations.

The reanalysis will be performed using the present Cray YMP 8 processors, 128 MW supercomputer, and the smaller Cray EL2. Other hardware includes a Robotic Silo, upgraded in August 1994 with 4490 STK drives, with storage capacity of 0.6 GB per tape. Over 2000 tape slots have been reserved for this project. Since the Cray YMP was saturated, the start

of the reanalysis had to wait for the new Cray C90 acquired by NMC to be installed (early 1994) and for the operational systems to migrate out of the Cray YMP (April 1994). Software used includes the UNICOS 7 operating system, NFS mount of Cray complex files, Bourne shell UNIX scripts, Fortran, some C, some X Windows, the Data Migration Facility, the Cray Reel Librarian, and the graphics system GrADS (COLA). Recent changes include the installation of UNICOS 8 and the replacement of the Bourne shell with the Korn (POSIX) shell. In addition, we expect that the Cray YMP and Cray EL2 will be replaced in 1995 by two Cray J916s.

b. Analysis scheme

The spectral statistical interpolation, a three-dimensional variational analysis scheme (Parrish and Derber 1992; Derber et al. 1991), is used as the analysis module. Its implementation in 1991 replacing an OI analysis led to major analysis and forecast improvements, especially in the Tropics, and a major reduction in the precipitation spinup. An important advantage of the SSI is that the balance imposed on the analysis is valid throughout the globe, thus making unnecessary the use of nonlinear normal-mode initialization. Recent enhancements, such as improved error statistics, and the use of the full tendency of the divergence equation in the cost function (replacing the original linear balance of the increments constraint), have also been included (Derber et al. 1994, Parrish et al. 1995). The SSI used in the reanalysis is the same as the system implemented in the operational system in January 1995, which was tested in parallel for over 10 months and resulted in significantly improved forecasts.

c. Model

The T62/28-level NCEP global spectral model is used in the assimilation system, as implemented in the NCEP operational system in December 1994. The vertical structure of the model is shown in Table 1. The model has five levels in the boundary layer and about seven levels above 100 hPa. The lowest model level is about 5 hPa from the surface, and the top level is at about 3 hPa. This vertical structure was chosen so that the boundary layer is reasonably well resolved and so that the stratospheric analysis at 10 hPa is not greatly affected by the top boundary conditions. The details of the model dynamics and physics are described in NOAA/NMC Development Division (1988), Kanamitsu (1989), and Kanamitsu et al. (1991). The model includes parameterizations of all major physical processes, that is, convection, large-scale precipitation, shallow convection, gravity wave drag, radiation with diurnal cycle and interaction with

TABLE 1. Model levels, midlevel sigma value, sigma thickness, geopotential thickness (m), and approximate location of the mandatory pressure levels (hPa).

Model level	Midlevel sigma	Delta sigma	Thickness	Mandatory pressure level
28	2.73	6.57	∞	3.0
27	10.06	7.29	5599	10.0
26	18.34	9.23	3828	20.0
25	28.75	11.60	3053	30.0
24	41.79	14.51	2621	
23	58.05	18.03	2342	50.0
22	78.15	22.22	2142	70.0
21	102.78	27.09	1984	100.0
20	132.61	32.62	1851	
19	168.23	38.67	1729	150.0
18	210.06	45.03	1612	200.0
17	258.23	51.35	1495	250.0
16	312.48	57.16	1376	300.0
15	372.05	61.97	1260	400.0
14	435.68	65.26	1139	
13	501.68	66.69	1017	500.0
12	568.09	66.06	895	
11	632.90	63.47	776	
10	694.26	59.19	664	700.0
9	750.76	53.72	560	
8	801.42	47.54	466	
7	845.79	41.15	384	850.0
6	883.84	34.93	313	
5	915.92	29.19	253	925.0
4	942.55	24.05	203	
3	964.37	19.59	162	
2	982.08	15.82	129	
1	995.00	10.00	80	1000.0

clouds, boundary layer physics, an interactive surface hydrology, and vertical and horizontal diffusion processes. A major difference in the model as described by Kanamitsu et al. (1991) is the use of a simplified Arakawa–Schubert convective parameterization scheme developed by Pan and Wu (1994) based on Grell (1993). Preimplementation experiments showed that the simplified Arakawa–Schubert scheme results in much better prediction of precipitation than the previous Kuo scheme over the continental United States, as measured by equitable threat scores over North America. In addition, the precipitation patterns over the Tropics are more realistic, with a smoother distribution and less concentration over tropical orographic features. Two other recent improvements were also implemented into the reanalysis model. The first is a better diagnostic cloud scheme (Campana et al. 1994), which has resulted in model-generated outgoing longwave radiation (OLR) in much better agreement with observations. The second is a new soil model, based on Pan and Mahrt (1987), which has also resulted in much more realistic surface temperature and more skillful predictions of precipitation over North America in the summer. These changes to the model were systematically tested by running 2 months of assimilations in summer and in winter, and 25 forecasts from each assimilation. Some tuning of the cloudiness and cloud optical properties were performed to correct systematic temperature and cloudiness errors. The final version of the model also produced good 5-day forecast scores.

d. Complex quality control of rawinsonde data

The method of CQC (Gandin 1988) is used to quality control the rawinsonde heights and temperatures. CQC first computes residuals from several independent checks (i.e., it computes the difference between an observation and the expected value for that observation from each check). It then uses these residuals together with an advanced DMA to accept, reject, or correct data (Collins and Gandin 1992). The checks included in the CQC code for rawinsonde heights and temperatures used for the reanalysis include: hydrostatic check, increment check with respect to the 6-h forecast, horizontal interpolation check, and vertical interpolation check. In addition, there is a baseline check based upon the difference between the station elevation and the elevation that is consistent with the reported surface pressure and the lowest two reported heights, using a standard lapse rate and the hydrostatic equation. Using the same information and assumptions, a mean sea level pressure may be obtained and compared with a forecast mean sea level pressure. In this way, both an increment and horizontal residual of mean sea level pressure are computed. The baseline check may allow the determination of errors in the location of stations as well as changes in their locations.

In addition to these checks, used operationally at NCEP, the reanalysis affords the possibility of also performing a temporal interpolation check, which cannot be done in the NWP system. The value of the heights and temperatures at observation time may be compared with those for 12 or 24 h earlier and later. The temporal residual is the difference between the reported height or temperature and the value interpolated from one value before and one after, when they are available. Statistics show this check to be of comparable value to the incremental check. It is used along with other available checks and is particularly useful in the data preprocessor, where the first guess and, hence, the incremental check are not available.

The CQC for rawinsonde heights and temperatures performs quite well. The code has been running operationally at NCEP for several years and has undergone steady improvement. At present, about 7% of the rawinsonde observations are found to have at least one error. Of the hydrostatically detectable errors in mandatory level heights and temperatures, 75% are confidently corrected, and 60% of the errors detected by use of the baseline check are also corrected. The absolute number of corrections for the early years of reanalysis may be anticipated to be smaller, depending upon data density, but there may be a higher percentage of data that need to be corrected.

The CQC methodology is also used to quality control the mandatory and significant level rawinsonde winds. We expect to develop a limited capability to correct some wind errors, for example, winds manually entered off by multiples of 100° in direction or multiples of 100 kt in speed.

e. Optimal interpolation quality control of all data

The OIQC (Woollen 1991; Woollen et al. 1994) was developed as the final screening for observations to be used in the data assimilation. The goal of OIQC is to detect and withhold from the assimilation data containing gross errors generated by instrumental, human, or communication-related mistakes that may occur during the process of making or transmitting observations. It also withholds observations with large errors of representativeness that are accurate but

whose measurements represent spatial and temporal scales impossible to resolve properly in the analysis-forecast system. The OIQC system uses the same statistical representativeness error model as the objective analysis system it precedes and, therefore, will detect observations unrepresentative for that system.

Three principles guide the OIQC algorithm: 1) use of multivariate three-dimensional statistical interpolation for obtaining comparison values for each observation from nearby neighbors, 2) a complex of independent quality control components consisting of interpolation and other types of checks that when evaluated collectively suggest whether errors exist in an observation, and 3) "nonhierarchical" decision making algorithm in which no final accept/reject decision is made for any data until all checks that may affect that decision are completed.

The OIQC components are interpolation checks; an OI of the appropriate variable made to each datum, from a group of observations nearby, forms a comparison value. Univariate and geostrophic horizontal checks are performed for each datum checked, as well as a univariate vertical (profile) check for sounding data. Temperature data is converted to units of equivalent thickness difference from a background (usually a 6-h forecast) for the checks, while wind data is checked in terms of vector wind deviations. A combination of individual check outcomes determines whether a datum is accepted by the system (see Kalnay et al. 1993; Woollen 1991; and Woollen et al. 1994 for further details).

For the reanalysis we plan to add to the complex of checks performed by the OIQC two more quantitative checks: one is produced by the time interpolation check of the CQC (see previous section), and the second is the deviation with respect to climatology, measured in units of the local analysis standard deviation climatology (see the discussion in section 4c). Both of these should be powerful additions to the QC and are made possible by the use of the BUFR with a QC system for storing data (see section 5f).

f. BUFR observation "events" files

The final step in observation preprocessing, described in section 3, consolidates incoming data from all sources into BUFR files, an internationally accepted standard format for level-2 data. Provisions have been made to archive in the reanalysis BUFR files, along with each original observation (in a "push-down" fashion), a spectrum of processing information, collectively known as "events." At present, these include an indication of the observation's source, all quality control decisions and their sources, a history of all modifications made to the observation prior to the analysis (QC corrections, radiation corrections, virtual temperature conversions, etc.), and various background quantities relevant to the analysis process (i.e., interpolated first-guess values, interpolated analyzed values, interpolated climatological values, interpolated climatological variances, and observation error estimates). As a result, the observational database produced by the reanalysis system contains a fairly complete processing history of each observation, which can be useful in the evaluation of the performance of the analysis procedures themselves, as well as to other reanalysis projects carried out at NCEP or at other research centers. The BUFR database archive is a major enhancement of prior NCEP observational formats and as such has been implemented as an operational product in the NCEP Global Data Assimilation System.

g. Optimal averaging

The NCEP/NCAR reanalysis system will include not only the computation of gridpoint (and spherical harmonics) values but also temporal and spatial averages over some prescribed areas. A new method, known as optimal averaging (Gandin 1993), will be used in the course of reanalysis. This method assures minimum (in statistical sense) root-mean-square averaging errors and, particularly important for the reanalysis purposes, it provides this estimated rms error as a by-product. The incorporation of OA will result in an increased ability for detection of climate change, because averaged values are less affected by the small-scale everyday variability that acts as noise and complicates the climate change signal detection.

In the reanalysis system, optimal averages over prespecified areas are computed for temperature, specific humidity, u and v components of the wind, and wind speed at seven (1000, 850, 700, 500, 300, 200, and 100 hPa) pressure levels. The horizontal areas currently include nine 20° latitude bands from the South Pole to the North Pole. The weights are computed by normalizing the optimal weights so that the sum of the weights over each area is equal to one. We also include the geographical regions chosen by the Intergovernmental Panel for Climate Change (IPCC) for climate change monitoring (IPCC 1990, p. 157), plus two regions covering South America: tropical South America, 10°N–20°S, 40°–80°W, and extratropical South America, 20°–50°S, 50°–70°W.

Two additional computations are included in the OA component. Optimal averages of the data increments (observation value minus first-guess value) are calculated by the same module that currently averages the actual observations. In addition, averages of the first guess over the same areas are computed directly from the spectral coefficients of the assimilation model.

h. Periodic forecasts from the reanalysis

J. M. Wallace suggested performing global forecasts during the reanalysis. In order to keep the computational requirement at a feasible level, we perform one 8-day forecast every 5 days. Such forecasts will be useful for predictability studies, indirect estimates of the accuracy of the analysis, and estimates of the impact of changes in the observing systems. They will also support the development of adaptive model output statistics from the long reanalysis (P. Dallavalle 1993, personal communication).

6. Reanalysis output

The design of the reanalysis output has been a major component of the project development [see also Schubert et al. (1993) for a discussion of the NASA reanalysis output]. During the April 1991 reanalysis workshop, it became clear that there are many different types of possible applications for the reanalysis output and that some of them (e.g., transport of greenhouse gases, which needs in principle all turbulent transports between any two layers) have storage requirements that far exceed what can be handled by the project. For this reason, it was decided that each unit of reanalysis output (1 mon) will include restart files, so that special purpose shorter reanalyses with extended output can be performed a posteriori.

The reanalysis gridded fields have been classified into four classes, depending on the relative influence of the observational data and the model on the gridded variable. An A indicates that the analysis variable is strongly influenced by observed data, and hence it is in the most reliable class (e.g., upper-air temperature and wind). The designation B indicates that, although there are observational data that directly affect the value of the variable, the model also has a very strong influence on the analysis value (e.g., humidity and surface temperature). The letter C indicates that there are no observations directly affecting the variable, so that it is derived solely from the model fields forced by the data assimilation to remain close to the atmosphere (e.g., clouds, precipitation, and surface fluxes). Finally, the letter D represents a field that is obtained from climatological values and does not depend on the model (e.g., plant resistance, land–sea mask). Appendix A contains the complete classification of variables. Although this classification is necessarily somewhat subjective, the user should exercise caution in interpreting the results of the reanalysis, especially for variables classified in categories B and C. In addition to this simple guidance, the user should keep in mind that quadratic variables (e.g., kinetic energy, transport of water vapor) are in general less reliable than the components from which they were computed. Appendix B contains the list of mandatory pressure, sigma, and isentropic levels of the output.

The reanalysis archive has been designed to satisfy two major requirements: 1) the output should be comprehensive, allowing, for example, the performance of detailed budget studies, and 2) it should be easily accessible to users interested in long time series of data. It became clear that it was not possible to satisfy both requirements with a single archival format. For this reason, the output module includes several different archives. Reanalysis information and selected output is also available on-line via the Internet (http//:nic.fb4.noaa.gov:8000). In this section we describe four types of archives and the automatic monitoring system that was designed to quality control the output.

a. BUFR observational archive

Reanalysis observational data undergo multiple processing stages, any of which may influence the quality of subsequent analysis and forecast products (see sections 4 and 5). For purposes of monitoring and review, and for research based on the reanalysis, it is useful to be able to trace the progression of QC and related processing to which any particular observation, or group of observations, has been subjected prior to its use (or nonuse) in the actual data assimilation. The BUFR observation event archive format (described in section 5f) has been designed to provide researchers with this capability.

Although the details of the BUFR format, and the BUFR structures devised to support the observation events archive, are rather complicated, a FORTRAN programmer interface package has been developed to simplify a user's interaction with these files and to enable fairly straightforward access to all of the archive information without the need for a great deal

of technical expertise in BUFR. These user-friendly FORTRAN interface routines, along with appropriate documentation and instructions for their use, will be available to reanalysis investigators.

b. Main synoptic archive

This is the most comprehensive archive of the reanalysis and will contain a large number of analysis and first-guess "pressure" fields at 0000, 0600, 1200, and 1800 UTC on a 2.5° latitude–longitude grid; "flux," "diagnostic," and "sigma" files on the model Gaussian grid; 192 × 94 points, in order to maintain maximum accuracy; and restart files at full resolution in order to ensure reproducibility. The complete list of output fields with their classification is given in appendix A. Table 2 summarizes the type and volume of the BUFR data archive and the gridded synoptic archives.

TABLE 2. Comprehensive "synoptic format" archives. See appendix B of the Kalnay et al. (1993) for a detailed list of fields, units, etc., contained in each of the files listed below.

File	Anal. (day^{-1})	Guess (day^{-1})	Total day^{-1}	MB $file^{-1}$	MB day^{-1}	GB mon^{-1}	GB $(40\ yr)^{-1}$
Restart files (non-GRIB)							
Sigma	4	4	8	1.9	15.4	0.47	225.6
Surface	4	4	8	1.2	9.6	0.29	139.2
GRIB SST/snow/sea ice							
SST	1	0	1	NA	0.09	0.00	1.24
Snow	1 $week^{-1}$	0	1 $week^{-1}$	NA	0.00	0.00	0.00
Sea ice	1	0	1	NA	0.00	0.00	0.01
BUFR observations							
Prepbufr	0	4	*5*	7.44	29.76	0.89	428.54
Fnlbufr	4	0	*10*	8.60	34.40	1.03	495.36
GRIB files (grid point)							
Pressure	4	4	8	1.73	13.84	0.42	199.30
Sigma	4	4	8	3.31	26.48	0.79	381.31
Grb2d	0	4	4	1.35	5.40	0.16	77.76
Grb3d	0	4	4	6.31	25.24	0.76	363.46
Isen	0	4	4	0.99	3.96	0.12	57.02
Non-GRIB files							
Zonal	4	4	8	0.05	0.42	0.01	6.11
Optavrg	4	0	4	0.03	0.12	0.00	1.72
Total (GB)						**4.94**	**2376.6**
STK cartridge (0.6 GB each)						**9**	**3960**

c. Reduced "time series" archive

This archive contains basic upper-air parameters on standard pressure levels (Table 3), selected surface flux fields (Table 4), and diabatic heating and radiation terms for each analysis cycle for the entire reanalysis period. Most of the data will be saved in GRIB format. The pressure level data will be saved on a 2.5° latitude–longitude grid, while the surface flux fields and radiation/diabatic heating data will be saved on a T62 Gaussian grid (192 × 94). In addition, monthly means of vorticity, divergence, virtual temperature, specific humidity, and surface pressure are saved in spherical harmonic form on sigma levels. Monthly means of the flux terms are stored on the Gaussian grid.

The radiation/diabatic heating data will be composed of two radiative terms (short- and longwave) and four diabatic heating terms (large-scale condensation, deep convection, shallow convection, and vertical diffusion). Monthly means of these data are stored at each sigma level of the 0000, 0600, 1200, and 1800 UTC cycles separately, so that the monthly mean diurnal cycle in these fields is preserved.

The data storage order will be markedly different from the manner in which model data have traditionally been stored at NCEP. Since the climate research community generally uses individual parameters at a single atmospheric level but for many time periods (rather that all parameters for a single time), much of the data are stored in chronological, not synoptic, order. That is, individual fields for a single atmospheric level are available for "all time" from a single data structure. The basic pressure level data and surface flux data will be stored in this order, which is referred to as "time series" in the rest of the paper.

d. Quick-look CD-ROM output

Following the suggestion of the advisory committee, we are creating a "quick look" database that can fit into a relatively small (1 yr^{-1}) number of CD-

TABLE 3. Data fields on standard pressure levels to be saved on a 2.5° lat × 2.5° long grid (144 × 73). The 17 levels used are 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 hPa.

Class	Field type
A	Zonal wind
A	Meridional wind
A	Geopotential height
A	Virtual temperature
A	Absolute vorticity
B	Vertical velocity (1000–100 hPa only)
B	Specific humidity (1000–300 hPa only)

ROMs. The complete content of the CD-ROMs is described in appendix C. It includes twice-daily values of *u, v, Z,* and *T* at three tropospheric and three stratospheric pressure levels (1000, 500, 200, 100, 50, 20 hPa). In addition, the CD-ROM contains daily values of total precipitable water, surface stress, latent and sensible heat flux, net long- and shortwave flux at the surface, precipitation, and surface pressure, SST, air surface temperature, soil temperature and moisture (two levels). It also includes isentropic potential vorticity; *u, v,* and *p* at three isentropic surfaces; monthly averages; and zonal cross sections of many fields and their covariances. See appendix C for the contents of the CD-ROMs.

A special CD-ROM, containing 13 years of selected observed, daily, monthly, and climatological data from the NCEP/NCAR Reanalysis, is included with this issue. See appendix E for a list of contents.

e. Automatic monitoring system for the reanalysis output

As previously noted, the NCEP reanalysis system was designed to perform 1 month of analyses (analyzed and archived every 6 h) every day, 5 days a week. Since the volume of reanalysis output is very large, it is not possible for a human monitor to review and check all the reanalysis products, detect major errors, drifts, etc. To fulfill this requirement, we developed an automatic monitoring system (Saha and Chelliah 1993; Kistler et al. 1994). At the end of each month of reanalysis we check the times series of geopotential height, zonal wind, meridional wind, temperature, and humidity at all standard pressure levels generated for every 6-h period (0000, 0600, 1200, and 1800 UTC).

TABLE 4. Surface flux data to be saved on the T62 Gaussian grid (192 × 94).

Class	Field type
B	Surface temperature
C	Skin temperature
B	2-m temperature
B	Surface pressure
D	Albedo
C	Surface sensitivity and latent fluxes
C	Top-of-the-atmosphere fluxes
B	Zonal wind at 10 m
B	Meridional wind at 10 m
C	Surface wind stress
A	Mean sea level pressure
B	Precipitable water
C	Snow depth
B	Snow cover
C	Precipitation (total and convective)
B	Mean relative humidity (multiple layers)
C	Soil wetness and temperature
C	Surface runoff
C	Cloud fraction (high, middle, low)
C	Cloud forcing, clear-sky fluxes
C	Gravity wave drag
B	Max and min temperature

To monitor the pressure time series, we use a preliminary climatology based on NCEP's GDAS over a 7-yr period from 1 July 1986 to 30 June 1993. From these daily values we computed monthly averages, standard deviations from the monthly means, standard deviations of the tendency (difference between successive analyses at 0000 UTC) and standard deviations of the interpolation check (difference between the analysis and the interpolation from analyses made 24 h before and after). These statistics were computed for geopotential height, zonal wind, meridional wind, and temperature at each grid point at 12 mandatory pressure surfaces and for humidity at 6 pressure surfaces. For monitoring the surface flux quantities, a 1-yr preliminary climatology (1 February 1992–31 January 1993) of daily surface flux files from a T62 model-based operational GDAS system at NCEP has also been created, since no long-term archive was available. These short-term preliminary climatologies will be later replaced by longer ones derived from the reanalysis itself.

We use the monthly statistics of climatological means, daily standard deviations, and daily standard deviations of the time interpolation check, at each 3D grid point and for each month of the year, to perform several statistical checks: we check for "field outliers" by computing for each variable the percentage of points whose distance to the climatological monthly mean is larger than two standard deviations. If this percentage is larger than the largest value observed in the 7-yr climatology, we identify the field as an outlier and proceed with further checks and diagnostics. These percentages are also graphically presented to the human monitor so that trends or jumps are immediately apparent. In addition, we also identify "single gridpoint outliers" that differ from the climate mean at each point by more than a specified number of standard deviations. This check has proven to be very important in identifying bad input data, such as radiosonde data or satellite wind data. A "time interpolation" check, is performed by computing similar statistics for the percentage of the points whose difference with a time interpolated field (from plus and minus 24 h values) is larger than expected.

Both of the above checks have proven to be very effective and have been consequently adapted for use in the data preprocessor (section 4). A number of subtle errors due to format changes, unusual data types, or quality-control decisions were discovered not only in the reanalysis system but also in the operational system. The usefulness of the check is due to the fact that the climatological standard deviations were computed for every month and for each grid point in 3D and were therefore much more sensitive than any other "gross check" previously used at NCEP. Further refinements to the automatic monitoring system will be developed using the first 5-yr reanalysis "climatology" now available.

With respect to the level-2 data, the human monitor will also have available the following information: output of the climatological check of the observations (available from the preprocessor before the monthly reanalysis), the normal operational output of the OIQC and the CQC, and several additional plots. These include a plot of the mean and rms data fits to the first guess and to the analysis, classified by region; a "curtain" time plot of the daily normalized rms fits of the data at all levels; and a plot of the data tossed out by the OIQC. These plots should allow a human monitor to check large amounts of data in order to detect serious problems.

7. The Climate Data Assimilation System

The reanalysis project originated with the idea of performing a "post analysis" with a CDAS, which would remain frozen into the future. In 1990, M. Cane and J. Nogués-Paegle of the advisory committee suggested that a very long reanalysis would be more useful than the CDAS alone. The development of the reanalysis system was then started and became the largest component of the project. It is clear that the combination of reanalysis for the past and the CDAS into the future, both using the same frozen system, will be much more helpful to researchers than either component alone.

The CDAS analysis will be performed within 3 days of the end of the month, with the same software as the reanalysis. This will allow for time to capture the bulk of any delayed data and serve as the basis for the generation of the monthly *Climate Diagnostics Bulletin* of CPC.

As noted before, our plans include a second phase of the CDAS/reanalysis to start sometime in 1998, after the first phase is completed. In the second phase, the reanalysis-2 will be performed with a 1999 state-of-the-art system, coupled with a corresponding CDAS-2, into the future. Such reanalysis would then be repeated every 5 years or so using the most advanced systems and the additional recovered data from the past. The CDAS-1, however, will be continued into the foreseeable future in order to maintain the longest homogeneous data assimilation product possible. Given that the CDAS-1 will become less expensive with time, it may be feasible to consider running a fixed observation system [choice (a) at the end of section 2] for comparison with the current reanalysis, which has considerable variations in the observing systems.

8. Coupling with the ocean

In the first phase of reanalysis we will couple the atmospheric analysis with the optimal interpolation reanalysis of SST for 1982 onward. For the earlier periods we will use the GISST data that the UKMO has offered to make available (Parker et al. 1993). The UKMO GISST analysis has recently been upgraded using EOFs, in collaboration with NCEP. In addition, a one-way coupled ocean reanalysis will be also performed.

a. The NCEP SST analysis

The NCEP routinely produces a 1° gridded SST

analysis using OI. The analysis is produced both daily and weekly, using 7 days of in situ data (ship and buoy) and bias-corrected satellite SST data. The first guess for the SST analysis is the preceding analysis. Because timescales of SST anomalies are of the order of months, the analysis from the previous week is a much better estimate of the current SST than climatology. There is a large-scale bias correction for satellite data, found necessary from past experience and because the OI method assumes that the data are unbiased (Reynolds 1988, 1993; Reynolds and Marsico 1993). The present version of the OI with the bias correction is a significant improvement over the earlier NCEP analysis and over any other analysis that uses uncorrected satellite data. In the Tropics, the equatorial eastern Pacific and Atlantic cold tongues are more realistically shown in the OI. At higher latitudes, the OI shows tighter gradients in the Gulf Stream, the Kuroshio, and the Falklands/Malvinas current regions. The statistics estimated in the process of developing the SST OI analysis show that ship SST observations have larger errors (1.3°C) compared to the errors of buoy and satellite SSTs (0.3°–0.5°C). In addition, the *e*-folding correlation scales have been found to range between 500 and 1200 km (Reynolds and Smith 1994).

The weekly version of the OI SST reanalysis has been computed for the period from November 1981 to the present. It is not practical to extend the period prior to November 1981 because the present operational satellite instrument [Advanced Very High Resolution Radiometer (AVHRR)] first became operational at that time. To develop a method to produce reliable SST analyses before November 1981, empirical orthogonal functions were computed from the monthly OI analyses for the 12-yr period from January 1982 through December 1993. A reduced set of spatial EOFs were then used as basis functions that were fitted to the in situ analyses to determine the correct temporal weighting of each function. Monthly SST anomalies were reconstructed from the spatial EOFs, from the temporal weights for the period 1950–81, and from a 2° grid from 45°S to 69°N. These fields capture most of the variance shown by in situ analyses while eliminating much of the noise due to sparse in situ data sampling. In a collaborative effort, the UKMO is testing a modification of the original GISST including the EOF expansion in order to enhance the signal-to-noise ratio for the periods before 1982 (N. Rayner 1995, personal communication).

b. The ocean reanalysis system

The coupled model project at NCEP has been performing 4D ocean analyses for the last 7 years in order to document more thoroughly current and past climate variability. These analyses also serve as the initial conditions and verification fields for the coupled ocean–atmosphere model used for multiseason forecasting (Ji et al. 1994). Since most of the potential extended predictability is thought to be the result of coupled interactions in the Tropics, the focus at NCEP has been the development of the ocean analysis in this region, but we plan to extend the ocean analysis domain to the entire globe.

The ocean model used was developed at GFDL. The Pacific model has a domain that extends from 45°S to 55°N and from 120°E to 70°W. The Atlantic domain extends from 100°W to 20°E and 50°S to 65°N. The bottom topography is variable, and there are 28 levels in the vertical. The zonal grid spacing is 1.5° in the Pacific and 1° in the Atlantic. The meridional grid spacing is one-third of a degree within 10° of the equator and gradually increases outside this zone to 1° poleward of 20°. Within 10° of the northern and southern boundaries the model fields are relaxed to climatological estimates. A Richardson number formulation for the vertical mixing is used in the upper ocean. Lateral mixing is formulated as proportional to the square of the equivalent horizontal wavenumber.

The data assimilation system (Derber and Rosatti 1989) is a 3D variational technique applied continuously in time. Presently, only thermal data is used in the analyses. All available temperature data from ships, satellite estimates, drifting and moored buoys, and expendable bathythermographs (XBTs) are used. Extensive quality control procedures have been developed to screen the data before they are used. Corrections are made to the model thermal fields in the upper ocean down to 720 m. This depth range contains the maximum depth of the bulk of the available subsurface thermal data from the T4- and T7-type XBT probes. Surface observations are kept in the analyses for 2 weeks, and subsurface observations are kept in for 4 weeks with weights varying linearly in time. Maximum weight is given at the observation time, with minimum weights at the beginning and end of the time interval when the data is being used.

Routine weekly ocean-model-based analyses are performed for the Pacific and Atlantic basins, with a 2-week delay in order to allow to be able to use the XBT data in the assimilation for 4 weeks. The mod-

els are forced with a weekly averaged stress field derived from the four-times-per-day near-surface winds produced by NCEP's global atmospheric analyses. These winds are converted to stress using a constant drag coefficient of 1.3×10^{-3}. The net heat flux used to force the model is set to zero in order to facilitate the evaluation of the heat fluxes in the analysis system. The net freshwater flux is also set to zero; once a year the salinity field is restored to the mean climatological field of Levitus (1982).

Using this system, ocean reanalyses have been performed for the Pacific and Atlantic basins. For this purpose, all available historical subsurface data were obtained from the archives and edited for the time period from June 1982 to the end of 1992. These were merged with the data that were available in real time. The Pacific reanalysis for June 1982–December 1992 has been completed, and a reanalysis for the Atlantic for the same period is underway.

Work on implementing a global-model-based ocean analysis system will start in 1995. The routine weekly analysis capability will be implemented first; the global reanalysis for the period 1982 to the present will be started in 1995, using the atmospheric reanalysis, and should take a few months for completion. With respect to earlier periods, the scarcity of subsurface ocean data implies that a meaningful reanalysis can only be done for the Northern Hemisphere, and only for the period starting in the late 1960s. Nevertheless, once a consistent set of forcing fields is available from the atmospheric reanalysis, we plan to perform the ocean reanalysis for the whole 40-year period (1957–96).

9. Preliminary results and reliability of the atmospheric reanalysis

The NCEP/NCAR reanalysis will produce 40 years of daily atmospheric and surface fields, which, for some variables, are close to a best estimate of the evolving state of the atmosphere. The analysis cycle, with the use of the 6-h forecast as a first guess, is able to transport information from data-rich to data-poor regions, so that even in relatively data-void areas the reanalysis can estimate the evolution of the atmosphere over both synoptic and climatological timescales.

A researcher using the reanalysis should be aware, however, that the different outputs are not uniformly reliable. As indicated in section 6, fields derived from a four-dimensional analysis are not equally influenced by observations. Some, such as upper-air mass and temperature fields (classified as A in appendix A) are generally well defined by the observations and, given the statistical interpolation of observations and first guess, provide an estimate of the state of the atmosphere better than would be obtained using observations alone. Others (classified as B) are partially defined by the observations but are also strongly influenced by the model characteristics. For example, the amount of moisture that the tropical model atmosphere can hold depends on its parameterization of cumulus convection, since some convection schemes tend to dry out the atmosphere more than others. Therefore, even if the analysis incorporates rawinsonde and satellite moisture data, the overall humidity will be influenced by the climatology of the model. This is even more true for quantities that are not directly observed or whose observations are not currently assimilated into the present analysis systems. Examples of these quantities (classified as C) are precipitation and surface fluxes. To the extent that the model and its physical parameterizations are realistic, these fields can be reliable and provide estimates as accurate as any other available, even on a daily timescale. However, they will have regional biases if the model tends to be biased. For example, over the southeastern United States, the model tends to be colder and drier than the atmosphere during the summer months. As a result, during the 6-h forecast in the analysis cycle, the model tends to precipitate the increment of moisture added by the rawinsonde observations during the analysis. This process of permanent spindown within the analysis cycle leads to excessive reanalysis precipitation in this area.

In this section we present a few results from the first 5 years of reanalysis (1985–89), comparisons with the then-operational GDAS, and several diagnostic studies. The impact that unavoidable changes in the observing systems (especially the introduction of new satellite data) will have on the reanalysis is also assessed. These results should provide an indication of the reliability of different reanalysis fields. Further results are presented in the proceedings of the Climate Diagnostics Workshop (Chelliah 1994; Saha et al. 1994; Janowiak 1994; Smith 1994; White 1994).

a. Global energy and water balance

Output from the NCEP/NCAR reanalysis includes many diagnostics of the physical forcing of the atmospheric flow, including complete surface energy and

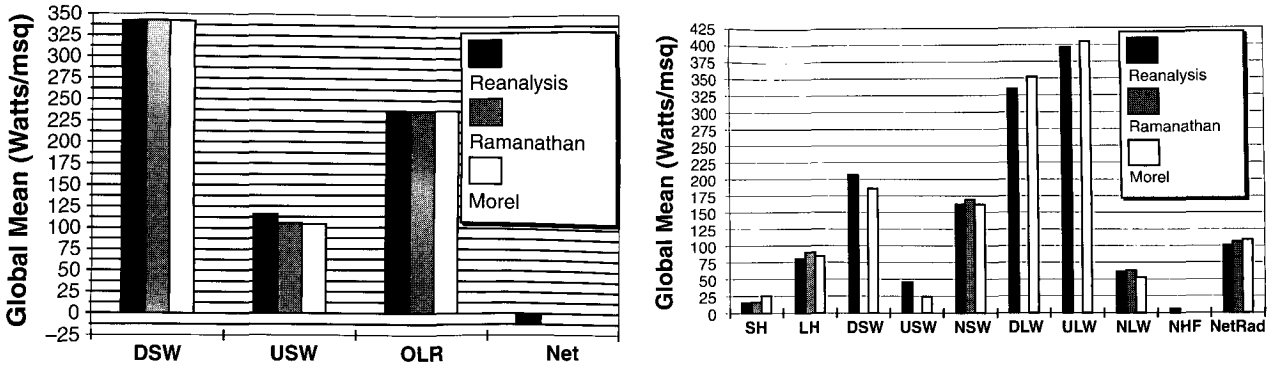


FIG. 4. (a) Global-mean radiative balance at the top of the atmosphere, and (b) surface energy budget for 1985–91 from reanalysis. Here, SH is sensible heat, LH is latent heat, SW is shortwave radiation, LW is longwave radiation, OLR is outgoing longwave radiation, D is downward, U is upward, N is net, HF is heat flux, RF is radiative flux, and Rad is radiation. Values from reanalysis are compared to two climatological estimates: Ramanathan et al. (1989) and Morel (1994).

hydrological budgets, the top-of-the-atmosphere radiation budget, angular momentum budgets, and monthly mean diabatic heating (White 1994). Figure 4a compares the global-mean radiation budget at the top of the atmosphere and Fig. 4b compares the surface energy budget from the reanalysis for 1985–91 with climatological estimates from Ramanathan et al. (1989) and Morel (1994). For most of these fields, classified as C, the reanalysis agrees with the climatologies as well as the different climatologies agree with each other. At the top of the atmosphere, upward shortwave radiation from reanalysis appears to be 11 W m^{-2} stronger than the climatological estimates (which are forced to be in balance), and the atmosphere loses 11 W m^{-2} to space. There is some evidence that the ocean surface albedo in the NCEP model is too high, and this may increase the upward solar radiation. At the surface the net radiation is 5-8 W m^{-2} less than the climatological estimates, and the atmosphere loses 5.5 W m^{-2} to the surface. Consistent with the loss of energy to space and the surface, the NCEP model cools slightly during the 6-h first-guess forecast. The zonal-mean and regional distributions of surface fluxes in reanalysis also appear to be consistent with climatological estimates.

Figure 5 displays monthly means of the global-mean hydrological budget and 12-mon running means of net atmospheric fluxes from the reanalysis for 1985–91. Over the entire period, evaporation exceeded precipitation by 0.04 mm day^{-1}. An annual cycle can be seen, with maximum values in July. Global-mean precipitation is within the range of climatological estimates. There is little evidence of any long-term drift in global averages in reanalysis.

b. The effect of SSM/I wind speeds

SSM/I data became available from the Defense Meteorological Satellite System (DMSP) in July 1987, and

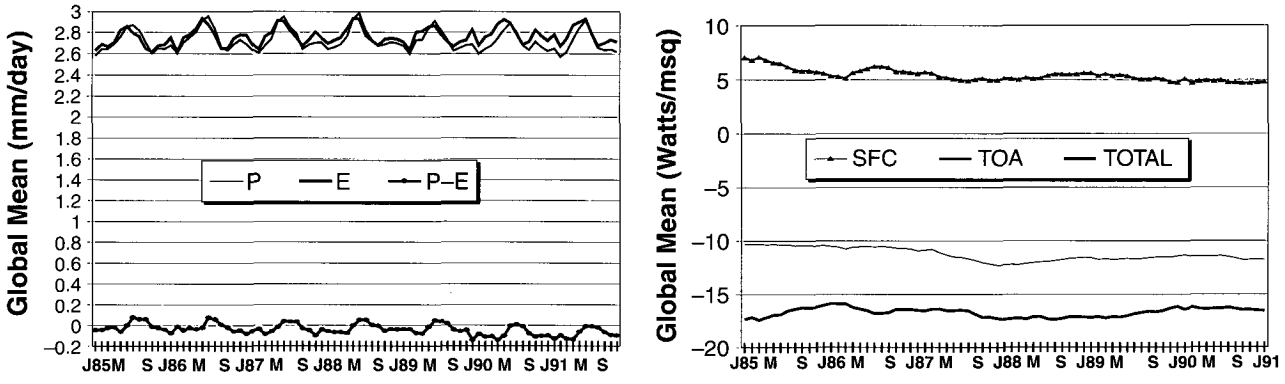


FIG. 5. (a) Monthly mean globally averaged precipitation (P) and evaporation (E), and their difference; (b) 12-mon running means of the net surface heat flux (SFC), net radiative flux at the top of the atmosphere (TOA), and the net flux out of the atmosphere (TOTAL) for 1985–91 from reanalysis.

NCEP started to use operationally ocean surface wind speeds derived with the algorithm of Goodberlet et al. (1989) in July of 1993, after parallel tests showed a positive impact of this data. We originally reprocessed a database of SSM/I radiances archived for climate purposes provided by NESDIS (N. Grody and R. Ferraro 1994, personal communication) to derive estimates of wind speeds. We used a neural network algorithm developed by Krasnopolsky et al. (1995). The neural network algorithm, which is nonlinear, results in significantly closer collocations with buoys than the previous operational algorithm of Goodberlet et al. (1989) and is less sensitive to clouds and moisture, giving a much larger coverage.

It was discovered in March 1995 that the SSM/I wind speeds assimilated for the period July 1987– December 1991, computed from the climate SSM/I database, did not contain a transformation from "antenna" temperature to "brightness" temperature. A preliminary evaluation estimated that this error created 10-m wind speeds with a positive bias of about 2 m s^{-1}. This bias resulted in an increase in surface fluxes of 5%–10%. With the corrected brightness temperature SSM/I data, the jump is much smaller. However, the very large volume of the original SSM/I radiance data, and even of the reduced SSM/I radiance data archived and quality controlled by F. Wentz (1994, NASA, personal communication), results in a very significant slowdown of the speed of the reanalysis processing. For this reason we have decided not to use SSM/I winds in the first phase of the reanalysis. The second phase will include the use of all SSM/I-derived products.

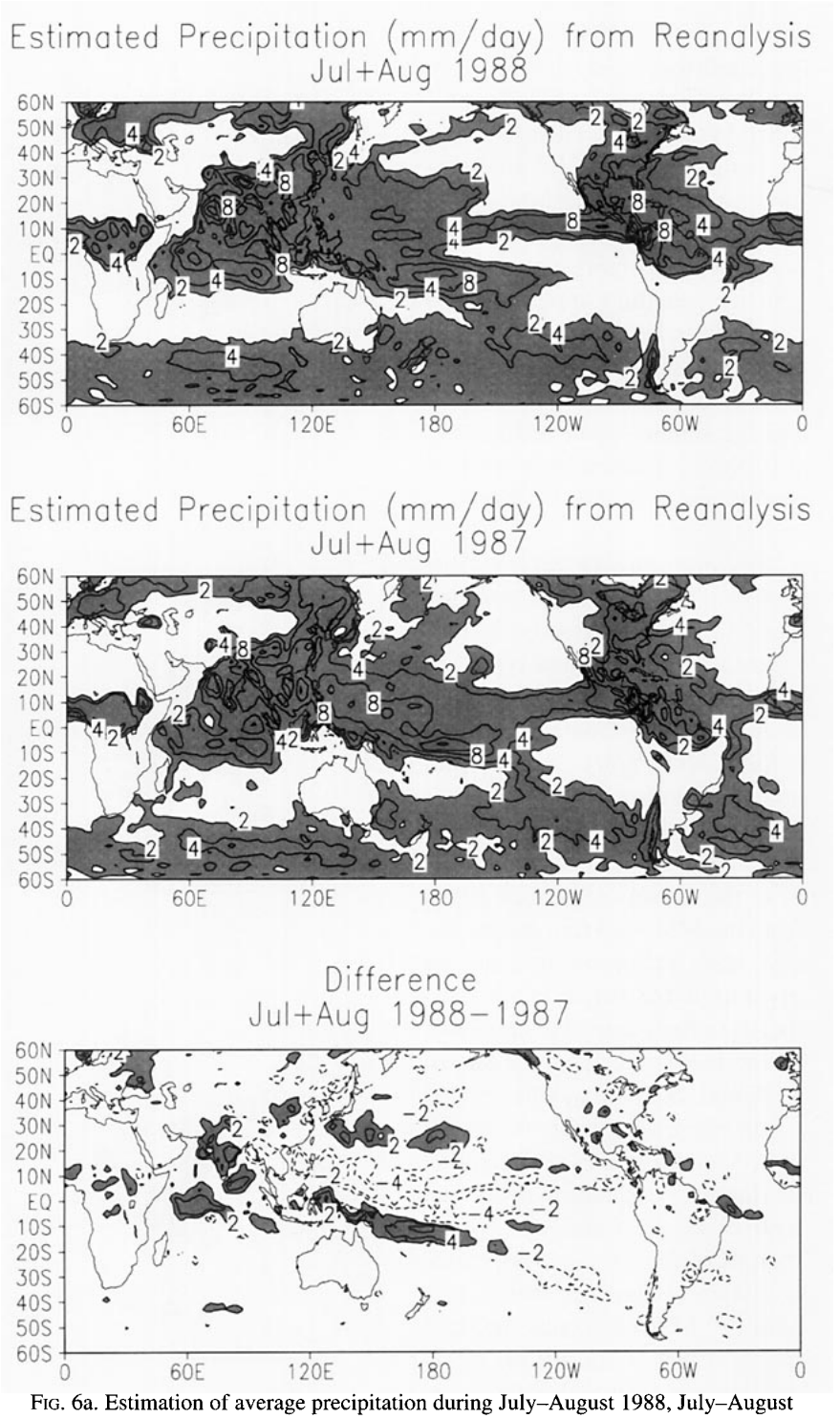


FIG. 6a. Estimation of average precipitation during July–August 1988, July–August 1987, and their difference, accumulated in the 6-h forecasts of the reanalysis. Contour lines at 2, 4, 8, 12, and 16 mm day^{-1}.

c. Sensitivity of monthly means

Before the reanalysis began, the impact on monthly mean fields of changes in the model used for the first guess was examined. These changes included the effect of horizontal and vertical resolution (T62 and T126) and different convection schemes. Salstein (1993) also examined the effect of horizontal resolu-

tion during May 1992. The results indicated that upper-level divergent flow, precipitation, and stratospheric winds were most sensitive to changes in the NCEP analysis-forecast system. The large-scale pattern of upper-level divergent flow (the scales represented by the velocity potential) appeared to be fairly robust in the Tropics; however, the magnitude of the upper-level divergent flow in the Tropics and the smaller-scale features are still poorly defined by a modern state-of-the-art analysis system.

d. Precipitation and soil moisture

Precipitation and soil moisture have C classifications, which means that data of these types are not assimilated but rather are derived completely from the model 6-h forecasts. Figures 6a,b depict precipitation from both reanalysis and from a dataset containing satellite-derived rainfall estimates over the oceans (Spencer 1993) from the MSU and rain gauge data over land. Precipitation maps are presented for the July–August means of the years 1987 and 1988, and soil moisture maps for August 1987 and 1988 are included.

We chose to compare the results between 1987 and 1988 because of the large precipitation shifts that were observed in many important regions of the Tropics, associated with the transition between "warm episode" El Niño–Southern Oscillation (ENSO) conditions (Rasmusson and Carpenter 1983) and "cold episode" conditions (Shukla and Paolino 1983) in the tropical Pacific during the 1986–88 period. SSTs were more than 3°C higher over much of the tropical Pacific during the warm event compared to the cold event, which had a large impact on the pattern of tropical convection and subsequent latent heat release. Ropelewski and Halpert (1987, 1989) have shown that precipitation tends to be less (more) than normal over India and the surrounding ocean during warm (cold) episodes, and they provide evidence that the Pacific ITCZ is displaced southward during warm episode conditions relative to cold episode conditions. The precipitation patterns over India and the Pacific ITCZ region were subsequently documented for these

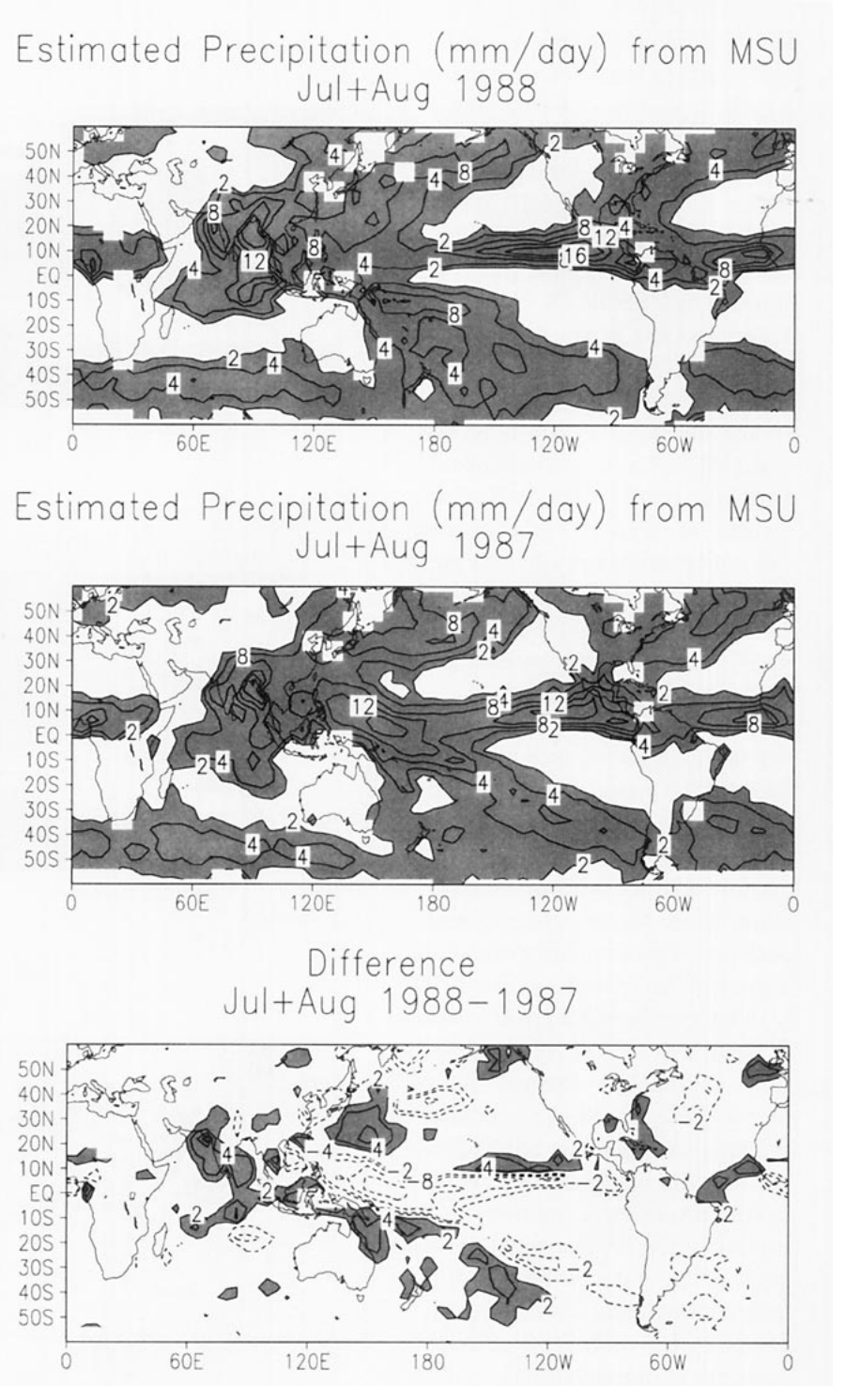


FIG. 6b. As in Fig. 6a but estimated from MSU and rain gauges.

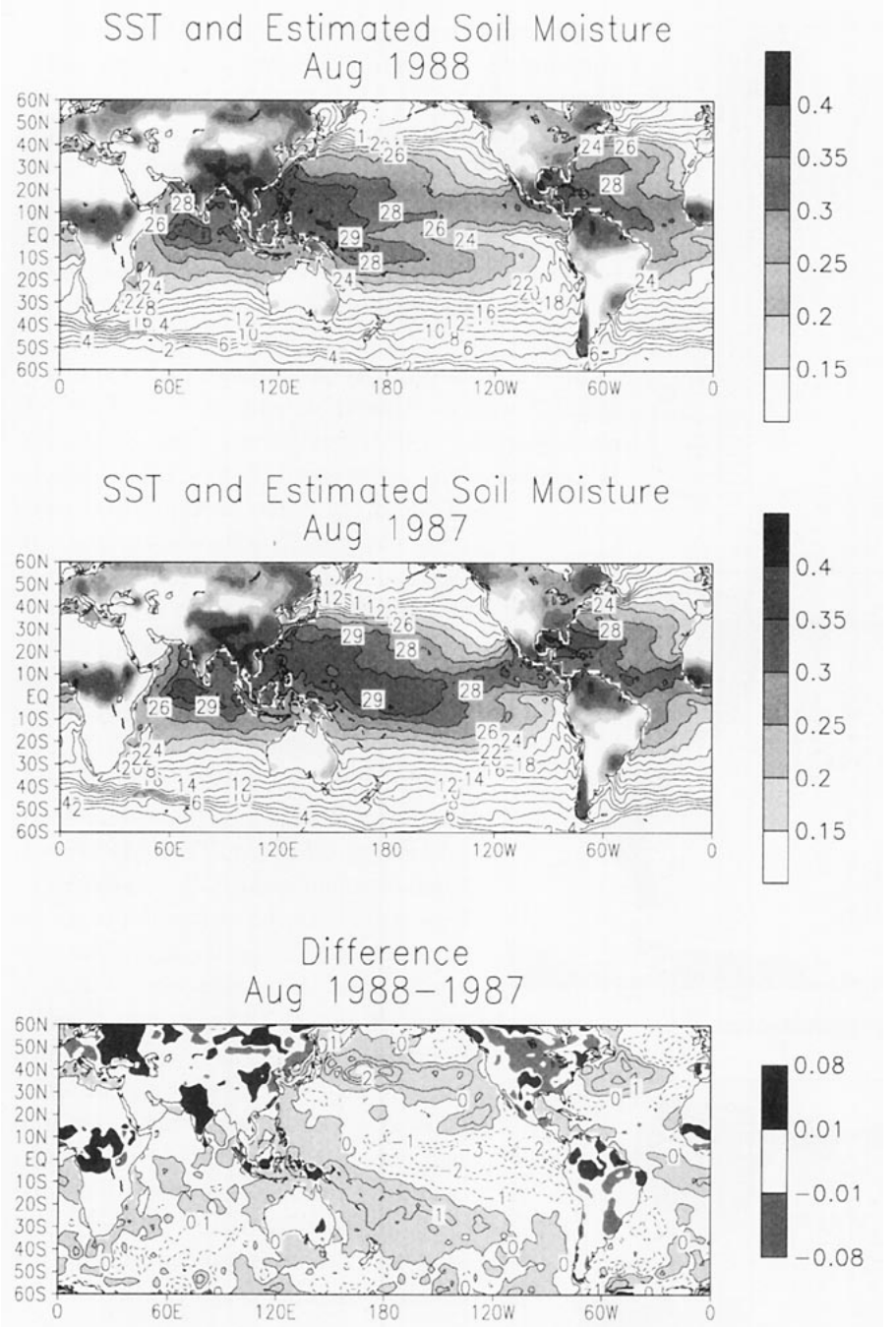


FIG. 6c. Soil moisture (relative to total field capacity, 200 cm) and sea surface temperatures averaged for August 1988, August 1987, and their difference.

specific events by Janowiak and Arkin (1991). The rainfall patterns between the Pacific warm and cold episodes that are represented in the MSU/rain gauge dataset described above are also consistent with the studies mentioned above.

The difference in the pattern of reanalysis precipitation between the northern summers of 1987 and 1988 compare well with those observed by the MSU/rain gauge dataset and with the studies mentioned above, over both the India region and the Pacific ITCZ. While the amplitude of the differences are considerably less than those of the MSU/rain gauge estimates in the tropical Pacific, there are also large differences in magnitude among independent satellite estimates of rainfall such as those based on MSU, infrared, and SSM/I data. The MSU estimates, like infrared algorithms, tend to overestimate rain rates and their geographical extent (P. Arkin 1995, personal communication). The reanalysis contains smaller scales than the MSU, which is probably too smooth.

The soil moisture changes (Fig. 6c) show that in the reanalysis India was wetter and North America was generally drier in 1988 than in 1987, as observed. The MSU/raingauge estimates (Fig. 6b) suggest that Central America had more rain in 1988 than in 1987 but that most of the rest of South America was drier in 1988. This is also true in the reanalysis but with weaker amplitudes. The reanalysis underestimates the intensity of the drought in the south and east of the United States during 1988.

Overall, the soil moisture generally appears to be reasonable and does not show a long-term tendency to drift into excessively dry or wet regimes (Fig. 7a), even without use of surface data but with a small nudging toward climatology. Figures 7b,c show the maximum and minimum monthly soil moisture content as a percentage of the field capacity (2.00 m) and the month of occurrence (only every third gridpoint arrow is plotted). As expected, the maximum soil moisture generally occurs at the end of the winter in midlatitudes and at the end of the monsoonal regime in the Tropics. The

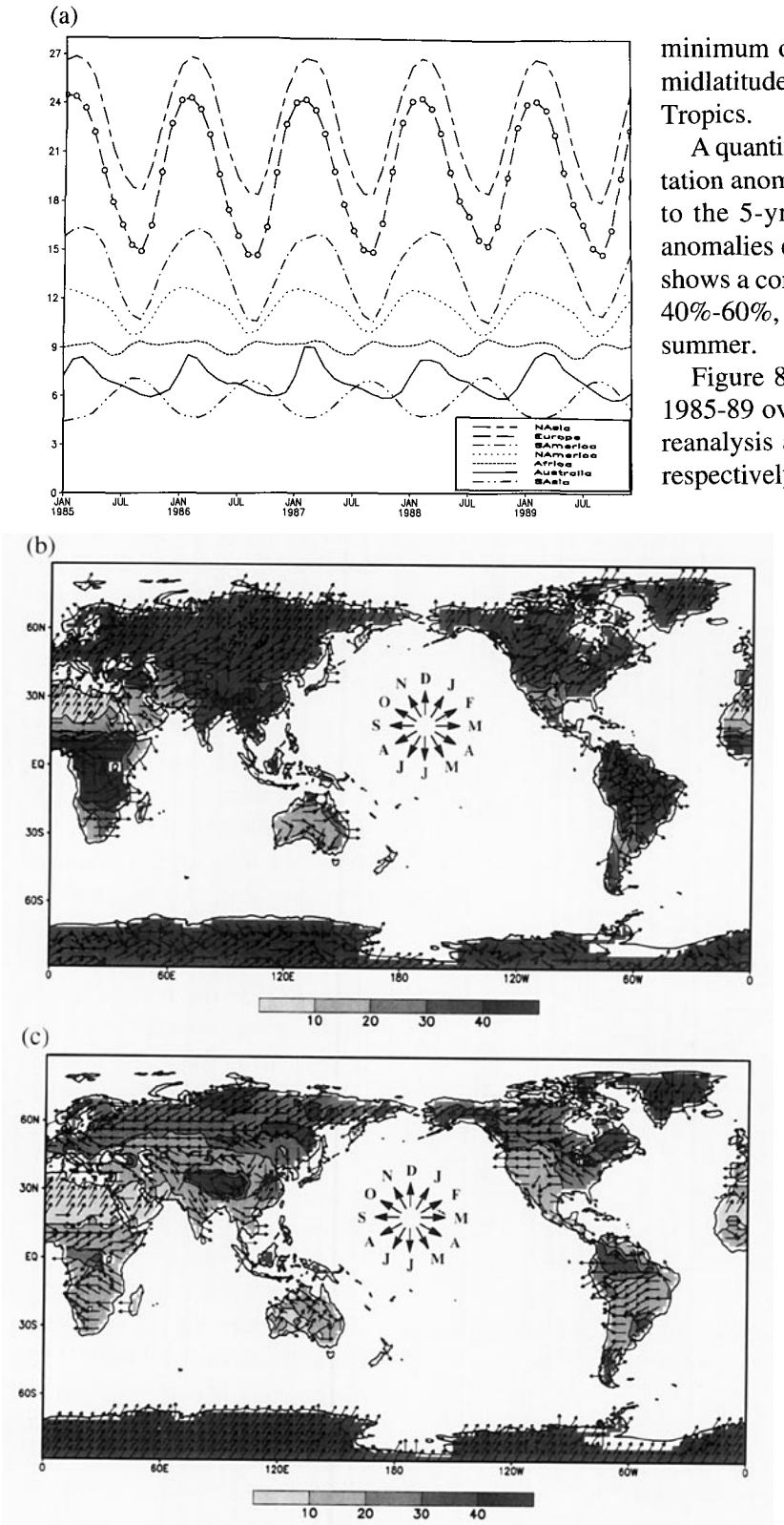


FIG. 7. (a) Evolution of the soil moisture content averaged over several regions of the world during the first 5 years of reanalysis. The units are percentages of the total field capacity (200 cm). (b) Average maximum soil moisture content estimated from 7 years of reanalysis. The arrows indicate the month at which the soil moisture is, on the average, maximum. (c) As in (b) but for the minimum soil moisture content.

minimum occurs generally after the summer in the midlatitudes and before the monsoon season in the Tropics.

A quantitative comparison of the reanalysis precipitation anomalies over the United States (with respect to the 5-yr mean) with the monthly precipitation anomalies estimated by the NCDC climate divisions shows a correlation pattern of the anomalies of about 40%-60%, somewhat higher in the winter than in the summer.

Figure 8 shows daily precipitation rates for May 1985-89 over the United States in the NCEP/NCAR reanalysis and in the observations, Figs. 8a and 8b, respectively. The corresponding standard deviations of the daily mean precipitation rates within May 1985–89 are also shown for the reanalysis (Fig. 8c) and for the observations (Fig. 8d). The observations were obtained from the hourly precipitation database compiled by the Techniques Development Laboratory of the National Weather Service (NWS) and contain about 300 NWS sites and 2500 cooperative stations. This data were gridded on a 2.5° grid (Y. Li 1994, personal communication). A comparison of Figs. 8a and 8b shows that in the southeastern United States the reanalysis precipitation is larger than observed by a factor of almost 2. As previously mentioned, this is due to a regional spindown of the model, which, being slightly drier and colder than the atmosphere, tends to rain out increments of moisture reintroduced by the analysis. However, the daily variability of the precipitation analysis compares quite well with the station variability (Figs. 8c,d).

e. Quasi-biennial oscillation and the stratospheric analysis

The operational NCEP global data assimilation system had poor resolution in the stratosphere until July 1993, when the vertical resolution was increased from 18 to 28 levels, and the top model levels was moved up to 2.7 hPa, changes that were also incorporated into the reanalysis. Figure 9a shows a 50-hPa Hovmoeller diagram (longitude–time) of the zonal velocity at 5°N to 5°S for the

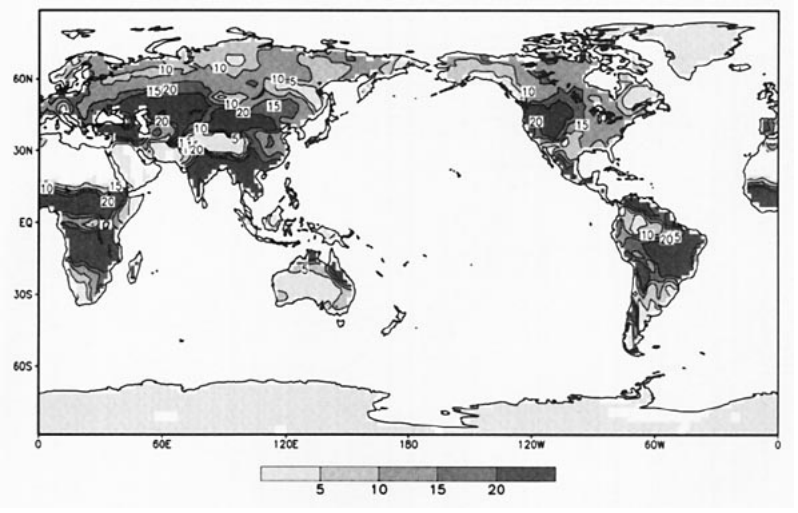


FIG. 7d. Difference between maximum and minimum soil moisture.

operational GDAS for the 5 years 1985–89 (denoted climate diagnostics data base or CDDB). This was the highest mandatory pressure level available in the GDAS at the time. Figure 9b shows the same 5-yr plot for the reanalysis. The quasi-biennial oscillation is very clear in the reanalysis and essentially absent in the operational analyses. Figure 10 shows a log pressure–time cross section from 100 to 10 hPa for a point in the reanalysis near Canton Island. The characteristics of the cross section are similar to those shown by Reed and Rogers (1962) for the Canton Island data. Their hand analysis of the station rawinsonde data also has a downward propagation of the phase, with a faster change from easterlies to westerlies than the reverse. Comparisons of the reanalysis near Singapore also show that there is good agreement with the rawinsonde data, indicating that the analysis system is able to assimilate well the data even in the upper-equatorial stratosphere.

More generally, comparisons of the stratospheric reanalysis with the off-line stratospheric analysis performed by NCEP (Finger et al. 1993) shows very good agreement (S.-K. Yang 1994, personal communication).

f. Impact of the FGGE observing system

A study was made to assess the impact that the introduction of the full satellite observing system will have on the reanalysis (Mo et al. 1995). Two sets of analyses and forecasts were made with and without

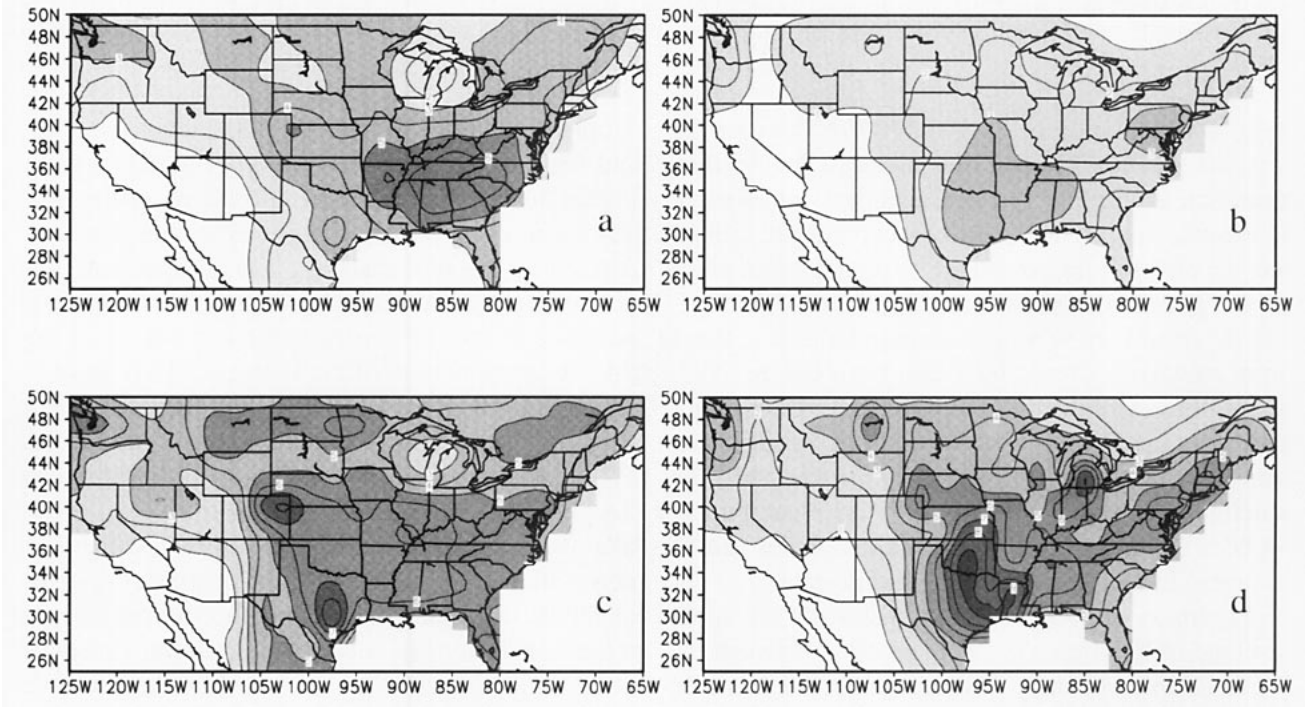


FIG. 8. Daily mean precipitation rates (mm day^{-1}) for May 1985–89 over the United States in (a) the NCEP/NCAR reanalysis and in (b) the observations. Standard deviation of the daily mean precipitation rates (mm day^{-1}) within May 1985–89 in (c) the NCEP/NCAR reanalysis and (d) the observations. Contour interval is 1 mm day^{-1}, greater than 1 mm day^{-1} is shaded.

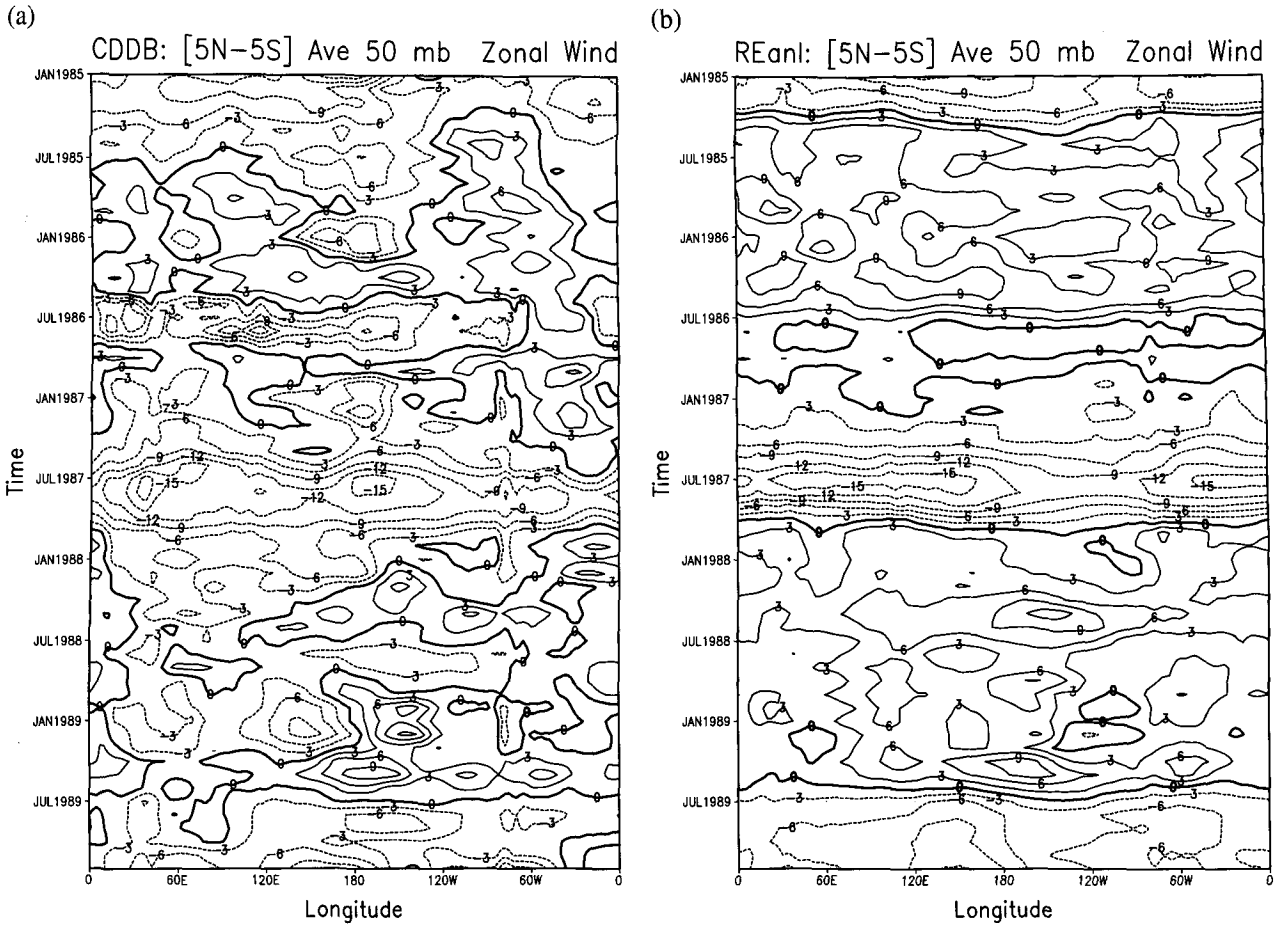


FIG. 9. a) Hovmoeller diagram (longitude–time) for the zonal wind component at 50 hPa at the equator (5°N–5°S), in the operational NCEP Global Data Assimilation System (from the climate diagnostics database). (b) As in (a) but for the reanalysis zonal winds.

the use of satellite data (SAT and NOSAT) within the data assimilation. The resulting impact is smaller than that obtained in previous satellite impact studies made using data from the FGGE (1979) experiment, reflecting the effect of improvements that have taken place in the global analysis scheme and the model. Overall, the results are very encouraging, indicating that a long reanalysis should be useful even before 1979, when the FGGE satellite-observing system was established: in the NH, the analyses of both primary variables and eddy fluxes are basically unaffected by the satellite data, and even in the SH a large component of both the monthly and the daily anomalies can be captured in the absence of the satellite data.

Figure 5a of Mo et al. (1995) showed the zonal average of the square of the correlation between the NOSAT and SAT daily analyses. It indicated that the NOSAT analysis explains close to 100% of the daily *variance* of the SAT geopotential height analysis in the NH extratropics, between 70% and 90% in the Tropics, more than 90% in the midlatitudes of the SH, and between 40% and 80% in the Antarctic region. Figure 5b showed the square correlation of the zonally asymmetric stationary (monthly averaged) eddies defined by the two analyses. The comparison suggested that NOSAT captures over 90% of the zonal *variance* of monthly mean stationary waves of the SAT analyses in most of the Tropics and SH down to 60°S, whereas in the NH extratropics the agreement is once again close to 100%. With respect to the bias of the zonally averaged values, the agreement between SAT and NOSAT is generally good, except above 200 hPa and in the polar regions. Obviously, the differences increase for more sensitive quantities, such as quadratic fluxes and their divergence. Typically, the relative differences between meridional fluxes of zonal momentum or heat estimated by the SAT and the NOSAT are less than 10% in the NH extratropics and less than 20% in the SH midlatitudes, but they can be as large as of order one in the Tropics, the strato-

sphere, and south of 60°S. Satellite data did not impact substantially the estimated precipitation fields.

g. Comparisons with other operational analyses

Finally, we compare the NCEP operational Global Data Assimilation System in use during 1992 (T126/18 levels and Kuo convection) with parallel runs using the new simplified Arakawa–Schubert scheme (Pan and Wu 1994) and the T62/28-level system adopted in the reanalysis. We also compared the NCEP and several other operational analyses. These differences are probably the best way to estimate the precision of the resulting analyses given similar observational databases and, therefore, are representative of the robustness of the NCEP reanalysis fields.

We define "internal analysis differences" as the rms difference between monthly means computed with NCEP systems using different models. "External analysis differences" are the rms differences between NCEP's monthly mean analysis and those of other operational systems.

The internal differences reflect the sensitivity to the first guess used in the analysis and are an estimate of the uncertainty in the monthly mean analysis of the NCEP system. For the Northern Hemisphere (20°–80°N), the internal differences are about 3 m at 850 hPa and 6 m at 500–200 hPa. In the Southern Hemisphere the internal differences are 5, 8, 15, and 30 m at 850, 500, 300, and 200, respectively, reflecting the much higher uncertainty introduced by the lack of rawinsonde data. The external differences between the NCEP analysis and the UKMO analysis are about 12, 7, 9, and 12 m at 850, 500, 300, and 200 hPa, respectively. The larger values at 850 hPa reflect the uncertainty introduced by different terrains and extrapolations below the surface. In the Southern Hemisphere (20°–80°S), the differences are about 20, 12, 15, and 25 m respectively. Comparisons with other operational systems were similar.

For the monthly mean wind analysis, in the NH the internal rms differences in both the zonal and meridional component are about 0.4 m s^{-1} at 850 hPa and 0.7 m s^{-1} at 200 hPa. The external rms differences are about 1 m s^{-1} at 850 hPa and 1.2 m s^{-1} at 200 hPa. In the Tropics (20°S–20°N), the internal rms differences for the zonal wind analysis are 0.7 m s^{-1} at 850 hPa and 2 m s^{-1} at 200 hPa. The external rms difference between the NCEP and the UKMO operational analysis in 1992 was 2 m s^{-1} at 850 hPa and 2.5 m s^{-1} at 200 hPa. For the meridional component, the rms differences were about 30% smaller. In the Southern

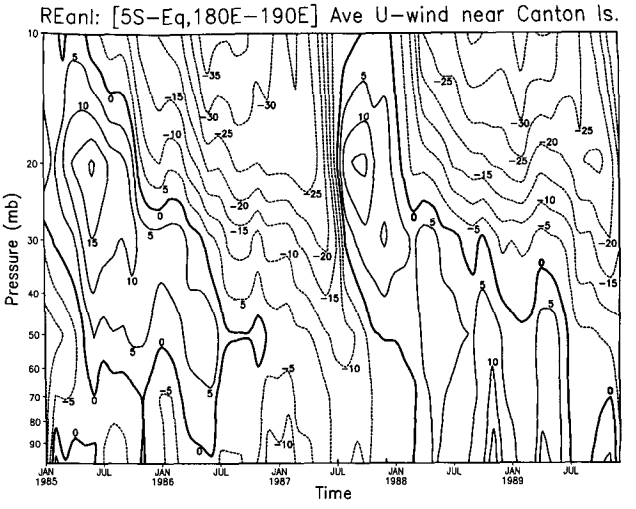


FIG. 10. Log pressure-time cross section of monthly zonal winds from the 1985–89 reanalysis for a point near Canton Island from 100 to 10 hPa.

Hemisphere, the internal differences at 850 hPa were about 0.8 m s^{-1}, the external differences were 1.3 m s^{-1}, and the meridional wind rms differences were about 20% smaller. At 200 hPa, both the internal and external rms differences were about 1.8 m s^{-1} for U and 1.2 m s^{-1} for V.

These figures indicate the precision with which modern analysis systems can determine monthly mean meteorological fields and can be regarded as a lower estimate of the accuracy with which such fields can be determined. Since all the analyses used employ very similar databases, it is likely that the true error is larger than the differences between the different analyses, since errors due to data gaps and measurement errors would be similar in the different analysis systems.

10. Summary

NCAR and NCEP have collaborated to create a very long reanalysis using a frozen, state-of-the-art global data assimilation system and a database as complete as possible. Changes in the observing systems can still produce perceived changes in the analyzed climate, but this problem is approached by producing parallel reanalyses (at least 1 year long) with and without using the new observing system for the period immediately after its introduction.

The system has been designed with advanced quality control and monitoring systems and can produce

1 month of reanalysis per "clock" day on a CRAY YMP/8 supercomputer. Different types of output archives are being created for different user needs, including a quick-look CD-ROM (one per year) archive with the most frequently used atmospheric fields, as well as surface, top-of-the-atmosphere, and isentropic fields.

The output variables have been classified into four classes, depending on the degree to which they are influenced by the observations and/or the model. Users are cautioned that C variables (such as precipitation and surface fluxes) are completely determined by the model, forced by the data assimilation to remain close to the atmosphere. Nevertheless, a comparison of these variables with different types of observations and climatologies show generally useful information on timescales from a few days to interannual variability.

Acknowledgments. We are very grateful to the Director of NCEP, Ron McPherson, for his continued enthusiastic support for this project. Without the generous and significant help provided by John Derber, David Parrish, Ken Campana, Robert Grumbine, Hua-lu Pan, Bert Katz, Jordan Alpert, Alan Basist, Mike Halpert, and Don Garrett from NCEP, Ki-Young Kim (KMA), and Steve Worley and Chi-Fan Shih from NCAR, this project could not have been carried out. George Murphy and Sarah Roy have supported our computer resources needs. The consultation with Cray Research analysts Ron Bagby, George Vanderberg, Mary McCann, and S. Wang has also been very helpful. David Parker and Nick Raynor (UKMO) have generously offered the pre-1982 sea surface temperature analysis. Brian Doty (COLA) created and enhanced the graphical package GrADS, which was crucial for this project, and Mike Fiorino (PCMDI) developed essential diagnostic software, including porting GrADS to the Cray system. Rex Gibson and Per Kallberg (ECMWF) and Siegfried Schubert and David Lamich (NASA/GLA) have provided us with observations to fill data gaps. Several countries, including Japan, China, Argentina, Brazil, and others have contributed data not available in real time on GTS especially for this project. Prof. E. Kung (University of Missouri) collaborated with the Chinese Meteorological Agency in recovering early Chinese rawinsonde data. The COADS (surface marine) dataset has many participants, including Scott Woodruff (NOAA/ERL), Steve Worley (NCAR), Joe Elms (NCDC), and Bob Keeley (Canada/Marine Environmental Data Service). Other key NCAR participants in data preparation are Bob Dattore (raobs, aircraft), Wilbur Spangler (upper air), Gregg Walters (upper air), Ilana Stern (surface), Roy Barnes (surface, TD13), and Joey Comeau (TOVS). Dick Davis (NCDC) has helped with surface data and consulting on old rawinsonde data, and John Lanzante (GFDL) with the preparation of old raob dataset TD54. People in many countries and laboratories also deserve credit for data preparation. Key participants are the observers around the world who work day and night to take the observations that make these analyses possible.

The project has been supported since its inception by the NOAA Office for Global Programs and by the National Weather Service (NCEP and GFDL), and by the National Science Foundation (NCAR). Without this generous support we would not have been able to develop this project. We are particularly grateful to Michael Coughlan and Ken Mooney for their support and guidance.

The advisory committee, chaired by Julia Nogués-Paegle from 1989 to 1993 and by Abraham Oort from 1993 to the present, has been a continuous source of advice and comfort when problems arose. Mark Cane, Julia Nogués-Paegle, and Milt Halem originally suggested performing a very long reanalysis, and J. Shukla spearheaded such a project throughout the research community. We are grateful to them and to the other members of the panel: Maurice Blackmon, Donald Johnson, Per Kallberg, David Salstein, Siegfried Schubert, John Lanzante, and James Hurrell.

Professor I. M. Navon and two anonymous reviewers made very helpful suggestions and corrected many errors in the original manuscript.

Appendix A: NCEP/NCAR reanalysis comprehensive output variables

The output variables are classified into four categories, depending on the relative influence of the observational data and the model on the gridded variable. An A indicates that the analysis variable is strongly influenced by observed data and, hence, it is in the most reliable class (e.g., upper-air temperature and wind). The designation B indicates that, although there are observational data that directly affects the value of the variable, the model also has a very strong influence on the analysis value (e.g., humidity and surface temperature). The letter C indicates that there are no observations directly affecting the variable, so that it is derived solely from the model fields forced by the data assimilation to remain close to the atmosphere (e.g., clouds and precipitation). Finally, the letter D represents a field that is fixed from climatological values and does not depend on the model (e.g., plant resistance, land–sea mask). This appendix contains the complete classification of variables. Although this classification is necessarily somewhat subjective, the user should exercise caution in interpreting the results of the reanalysis, especially for variables classified in categories B and C. In addition to this rule of thumb, the user should keep in mind that quadratic variables (e.g., kinetic energy, transport of water vapor) are in general less reliable than the components from which they were computed.

a. Standard GRIB output

1) PRESSURE: PRESSURE COORDINATE OUTPUT

Regular (2.5° lat × 2.5° long) grid. All fields are instantaneous values at a given time:

Class	Field type
A	Geopotential height (gpm) at 17 levels
A	u wind (m s^{-1}) at 17 levels
A	v wind (m s^{-1}) at 17 levels
A	Temperature (K) at 17 levels
B	Pressure vertical velocity (Pa s^{-1}) at 12 levels
B	Relative humidity (%) at 8 levels
A	Absolute vorticity (s^{-1}) at 17 levels
A	u wind of the lowest 30-hPa layer (m s^{-1})
A	v wind of the lowest 30-hPa layer (m s^{-1})
B	Temperature of the lowest 30-hPa layer (K)
B	Relative humidity of the lowest 30 hPa (%)
B	Pressure at the surface (Pa)
B	Precipitable water (kg m^{-2})
B	Relative humidity of the total atmospheric column (%)
A	Temperature at the tropopause (K)
A	Pressure at the tropopause (Pa)
A	u wind at the tropopause (m s^{-1})
A	v wind at the tropopause (m s^{-1})
A	Vertical speed shear at the tropopause (1 s^{-1})
B	Surface lifted index (K)
B	"Best" (4 layer) lifted index (K)
A	Temperature at the maximum wind level (K)
A	Pressure at the maximum wind level (Pa)
A	u wind at the maximum wind level (m s^{-1})
A	v wind at the maximum wind level (m s^{-1})
D	Geopotential height at the surface (gpm)
A	Pressure reduced to MSL (Pa)
B	Relative humidity in three sigma layers: 0.44–0.72, 0.72–0.94, 0.44–1.0 (%)
B	Potential temperature at the lowest sigma level (K)
B	Temperature at the lowest sigma level (K)
B	Pressure vertical velocity at the lowest sigma level (Pa s^{-1})
B	Relative humidity at the lowest sigma level (%)
B	u wind at the lowest sigma level (m s^{-1})
B	v wind at the lowest sigma level (m s^{-1})

2) GRB2D: TWO-DIMENSIONAL DIAGNOSTIC FILE

Class	Field type
C	Cloud forcing net longwave flux at the top of atmosphere (W m^{-2})
C	Cloud forcing net longwave flux at the surface (W m^{-2})
C	Cloud forcing net longwave flux for total atmospheric column (W m^{-2})
C	Cloud forcing net solar flux at the top of the atmosphere (W m^{-2})
C	Cloud forcing net solar flux at the surface (W m^{-2})
C	Cloud forcing net solar flux for total atmospheric column (W m^{-2})
C	Convective precipitation rate (kg m^{-2} s^{-1})
C	Clear sky downward longwave flux at the surface (W m^{-2})
C	Clear sky downward solar flux at the surface (W m^{-2})
C	Clear sky upward longwave flux at the top of the atmosphere (W m^{-2})
C	Clear sky upward solar flux at the top of atmosphere (W m^{-2})
C	Clear sky upward solar flux at the surface (W m^{-2})
C	Cloud work function (J Kg^{-1})
C	Downward longwave radiation flux at the surface (W m^{-2})
C	Downward solar radiation flux at the top of the atmosphere (W m^{-2})
C	Downward solar radiation flux at the surface (W m^{-2})
C	Ground heat flux (W m^{-2})
D	Ice concentration (ice = 1; no ice = 0) (1/0)
D	Land–sea mask (1 = land; 0 = sea) (integer)
C	Latent heat flux (W m^{-2})
C	Near IR beam downward solar flux at the surface (W m^{-2})
C	Near IR diffuse downward solar flux at the surface (W m^{-2})
C	Potential evaporation rate (W m^{-2})
C	Precipitation rate (kg m^{-2} s^{-1})
C	Pressure at high-cloud top (Pa)
C	Pressure at high-cloud base (Pa)
C	Pressure at middle-cloud top (Pa)
C	Pressure at middle-cloud base (Pa)
C	Pressure at low-cloud top (Pa)
C	Pressure at low-cloud base (Pa)
C	Pressure at the surface (Pa)
C	Runoff (kg m^{-2} per 6-h interval)
D	Surface roughness (m)
C	Nearby model level of high-cloud top (integer)
C	Nearby model level of high-cloud base (integer)
C	Nearby model level of middle-cloud top (integer)
C	Nearby model level of middle-cloud base (integer)

C	Nearby model level of low-cloud top (integer)
C	Nearby model level of low-cloud base (integer)
C	Sensible heat flux (W m^{-2})
C	Volumetric soil moisture content (fraction) (two layers)
B	Specific humidity at 2 m (kg kg^{-1})
C	Total cloud cover of high-cloud layer (%)
C	Total cloud cover of middle-cloud layer (%)
C	Total cloud cover of low-cloud layer (%)
B	Maximum temperature at 2 m (K)
B	Minimum temperature at 2 m (K)
B	Temperature at the surface (skin temperature) (K)
C	Temperature of the soil layer (three layers) (K)
B	Temperature at 2 m (K)
C	Temperature of high-cloud top (K)
C	Temperature of low-cloud top (K)
C	Temperature of middle-cloud top (K)
C	Zonal gravity wave stress (N m^{-2})
B	Zonal component of momentum flux (N m^{-2})
B	*u* wind at 10 m (m s^{-1})
C	Upward longwave radiation flux at the top of the atmosphere (W m^{-2})
C	Upward longwave radiation flux at the surface (W m^{-2})
C	Upward solar radiation flux at the top of the atmosphere (W m^{-2})
C	Upward solar radiation flux at the surface (W m^{-2})
C	Meridional gravity wave stress (N m^{-2})
C	Visible beam downward solar flux at the surface (W m^{-2})
C	Visible diffuse downward solar flux at the surface (W m^{-2})
C	Meridional component of momentum flux (N m^{-2})
B	*v* wind at 10 m (m s^{-1})
C	Water equivalent of accumulated snow depth (kg m^{-2})

3) GRB3D: THREE-DIMENSIONAL DIAGNOSTIC FILE

Gaussian grid (192 × 94) on 28 model levels. All fields are average of 6-h integration starting from a given time:

Class	Field type
C	*Deep convective* heating rate (K s^{-1})
C	Deep convective moistening rate [kg (kg s^{-1})$^{-1}$]
C	Large-scale condensation heating rate (K s^{-1})
C	Longwave radiative heating rate (K s^{-1})
C	Shallow convective heating rate (K s^{-1})
C	Shallow convective moistening rate [kg (kg s^{-1})$^{-1}$]
C	Solar radiative heating rate (K s^{-1})
C	Vertical diffusion heating rate (K s^{-1})
C	Vertical diffusion moistening rate [kg (kg s^{-1})$^{-1}$]
C	Vertical diffusion zonal acceleration [m (s s^{-1})$^{-1}$]
C	Vertical diffusion meridional acceleration [m (s s^{-1})$^{-1}$]

4) SIGMA

Gaussian grid (192 × 94) on 28 model levels or surface. All fields are instantaneous values at a specified time:

Class	Field type
A	Relative vorticity (28 levels) (s^{-1})
B	Divergence (28 levels) (s^{-1})
A	Temperature (28 levels) (K)
B	Specific humidity (28 levels) (kg kg^{-1})
A	*x* gradient of log pressure (surface) (m^{-1})
A	*y* gradient of log pressure (surface) (m^{-1})
A	*u* wind (28 levels) (m s^{-1})
A	*v* wind (28 levels) (m s^{-1})
A	Pressure (surface) (Pa)
A	Geopotential height (surface) (gpm)
A	*x* gradient of height (surface) (m m^{-1})
A	*y* gradient of height (surface) (m m^{-1})

5) ISEN: ISENTROPIC COORDINATE OUTPUT

Gaussian grid (192 × 94) most on 10 isentropic levels. All fields are instantaneous values at a specified time:

Class	Field type
A	Potential temperature (surface) (K)
A	Temperature (K)
A	*u* wind (m s^{-1})
A	*v* wind (m s^{-1})
B	Pressure vertical velocity (Pa s^{-1})
B	Relative humidity (%)
A	Montgomery stream function (m^2 s^{-1})
B	Brunt–Väisälä frequency squared (s^{-2})
B	Potential vorticity [m^2 (s kg^{-1})$^{-1}$]

b. Other non-GRIB output files

ZONAL DIAGNOSTIC FILE (BINARY)

Average over 90°–60°S, 60°–30°S, 30°S–30°N, 30°–60°N, 60°–90°N, and global. Unmarked fields are instantaneous values at a given time. "Av" indicates average during the 6-h integration:

Class	Field type
A	u component of wind (m s^{-1}) at 28 model levels
A	v component of wind (m s^{-1}) at 28 model levels
A	Virtual temperature (K) at 28 model levels
B	Specific humidity (g g^{-1}) at 28 model levels
B	Squared vorticity (1 s^{-2}) at 28 model levels
C	Squared divergence (1 s^{-2}) at 28 model levels
B	Pressure vertical velocity (Pa s^{-1}) at 28 model levels
A	Temperature (K) at 28 model levels
B	Relative humidity (%) at 28 model levels
B	Kinetic energy (m^2 s^{-2}) at 28 model levels
C	Convective heating (K s^{-1}) at 28 model levels (Av)
C	Large-scale heating (K s^{-1}) at 28 model levels (Av)
C	Shallow convection heating (K s^{-1}) at 28 model levels (Av)
C	Vertical diffusion heating (K s^{-1}) at 28 model levels (Av)
C	Convective moistening [g (g s^{-1})$^{-1}$] at 28 model levels (Av)
C	Shallow convection moistening [g (g s^{-1})$^{-1}$] at 28 model levels (Av)
C	Vertical diffusion moistening [g (g s^{-1})$^{-1}$] at 28 model levels (Av)
C	Zonal accel by vertical diffusion (m s^{-2}) at 28 model levels (Av)
C	Meridional acceleration by vertical diffusion (m s^{-2}) at 28 model levels
C	Shortwave radiation heating (K s^{-1}) at 28 model levels (Av)
C	Longwave radiation heating (K s^{-1}) at 28 model levels (Av)
C	Total precipitation (kg m^{-2}) (Av)
C	Convective precipitation (kg m^{-2}) (Av)
C	Sensible heat flux (w m^{-2}) (Av)
C	Latent heat flux (w m^{-2}) (Av)
B	Zonal stress (dyn m^{-2}) (Av)
B	Meridional stress (dyn m^{-2}) (Av)
C	Rain area coverage (%)
C	Convective rain area coverage (%)
B	Surface pressure (hPa)
C	Surface skin temperature (K)
C	Soil wetness (cm)
C	Snow depth (m)
C	10-cm-deep soil temperature (K)
C	50-cm-deep soil temperature (K)
D	500-cm-deep soil temperature (K)
C	Surface net shortwave flux (W m^{-2}) (Av)
C	Surface net longwave flux (W m^{-2}) (Av)
B	Relative humidity at the lowest model level (%)
B	Virtual temperature at the lowest model level (K)
B	Temperature at the lowest model level (K)
B	Specific humidity at the lowest model level (K)
D	Surface roughness (m)
D	Land–sea sea-ice mask (integer)
C	Zonal acceleration by gravity wave drag (m s^{-2}) (Av)
C	Meridional acceleration by gravity wave (m s^{-2}) (Av)
B	Surface torque [g (m^2 s^{-2})$^{-1}$] (Av)
C	Gravity wave drag torque [g (m^2 s^{-2})$^{-1}$] (Av)
B	Mountain torque [g (m^2 s^{-2})$^{-1}$] (Av)
B	Total angular momentum (m^2 s^{-1})
B	Planetary angular momentum (m^2 s^{-1})

c. Restart files (binary)

Spectral (28 model levels) or Gaussian grid (192 × 94). All fields are instantaneous values at a specified time.

1) SIGMA SPECTRAL COEFFICIENT FILE

Class	Field type
D	Surface geopotential
B	Natural log of surface pressure
A	Virtual temperature
B	Divergence
A	Vorticity
B	Specific humidity

2) SURFACE FILE (ON GAUSSIAN GRID)

Class	Field type
C	Earth surface temperature (K)
C	Soil moisture level 1 (% volume)
C	Soil moisture level 2 (% volume)
C	Snow depth (m)

C	Soil temperature level 1 (K)
C	Soil temperature level 2 (K)
C	Soil temperature level 3 (K)
D	Surface roughness length (m)
C	Convective cloud cover (%)
C	Convective cloud bottom height (sigma)
C	Convective cloud top height (sigma)
C	Albedo (fraction)
C	Snow/ice/land mask
D	Minimum stomatal resistance (s m^{-1})
C	Canopy water content (m)
C	Ratio of 10-m and lowest sigma level winds (fraction)

Appendix B: Output levels

Standard pressure levels (hPa):

1000	925	850	700	600	500	400	300	250
100	150	100	70	50	30	20	10	

Isentropic surfaces (K):

650	550	450	400	350	330	315	300	290	280	270

Sigma levels:

0.9950	0.9821	0.9644	0.9425	0.9159	0.8838
0.8458	0.8014	0.7508	0.6943	0.6329	0.5681
0.5017	0.4357	0.3720	0.3125	0.2582	0.2101
0.1682	0.1326	0.1028	0.0782	0.0580	0.0418
0.0288	0.0183	0.0101	0.0027		

Appendix C: NMC/NCAR reanalysis output on CD-ROM

Some of the reanalysis products will be distributed on CD-ROM. Currently, two types of CD-ROMs are being planned. The first would contain reanalysis products for a single year (one CD-ROM per year). The second type would be produced after about 10 years and would contain time series of relatively few variables. We believe this is an efficient way to satisfy the requirements of most members of the meteorological community, many of whom were consulted in the preparation of the output list. The following is the plan for the first type of CD-ROM.

Note that the output variables should be classified into four categories, depending on the relative influence of the observational data and the model on the gridded variable (see appendix A for a complete classification). The user should exercise caution in interpreting the results of the reanalysis, especially for variables classified into categories B and C.

0000 and 1200 UTC Analyses	**Estimated size**
u, v, temperature at 850, 500, 200 hPa	114 MB
Geopotential height 1000, 850, 700, 500, 300, 200 hPa (sea level pressure can be derived from the above fields)	29 MB
Surface pressure	21 MB
Omega at 500 mb	10 MB
Precipitable water	10 MB
Temperature at 2 m	18 MB
Specific humidity at 2 m	14 MB
u, v at 10 m	29 MB
RH at 500 and 200 hPa	13 MB
Total for 0000 and 1200 UTC analyses	**258 MB yr^{-1}**
Daily averaged analyses	
Zonal, meridional wind stress	19 MB
Net short/longwave flux at surface	14 MB
Precipitation	8 MB
Latent/sensible heat flux	16 MB
Model OLR	7 MB
Downward shortwave flux at surface	7 MB
Outgoing shortwave flux at top	7 MB
Tmin, Tmax (24-h period)	17 MB
Skin temperature (includes SST)	9 MB
Snow (liquid water equivalent)	13 MB
Total for daily averaged fields	**117 MB yr^{-1}**
0000 (isentropic) and 1200 (stratospheric) analyses	
Height and temperature at 100, 50 and 20 hPa	29 MB
u, v at 100, 50, and 20 hPa	28 MB
Potential vorticity on three θ surfaces (315, 330, 450 K)	15 MB
u, v on three theta surfaces (315, 330, 450 K)	30 MB

Temperature on three theta surfaces (315, 330, 450 K)	14 MB
Total for isentropic and stratospheric analyses	**117 MB yr^{-1}**
All cross sections (monthly averaged)	**2 MB yr^{-1}**
Monthly means, variances, and covariances	**138 MB yr^{-1}**
Observed OLR	**5 MB yr^{-1}**
GrADS control and index files	**24 MB yr^{-1}**

Documentation

Kalnay et al. (1993 with updates)

Office Note 388
(GRIB table of local definitions, documentation)

Miscellaneous

Total documentation volume	3 MB
Software to read grib (PC-GrADS, wgrib)	6 MB
Estimated total volume	**670 MB yr^{-1}**

Appendix D: List of acronyms

AVHRR	Advanced Very High Resolution Radiometer
BUFR	binary universal format representation
CDAS	Climate Data Assimilation System
CDDB	Climate diagnostics database
COADS	Comprehensive Ocean-Atmosphere Data Set
COLA	Center for Ocean-Land-Atmosphere
CPC	Climate Prediction Center (ex-CDC)
CQC	complex quality control
DMA	decision making algorithm
DOE	Department of Energy
DMSP	Defense Meteorological Satellite Program
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño-Southern Oscillation
ERL	Environmental Research Laboratories
FGGE	First GARP Global Experiment (1979)
FTP	file transfer protocol
GARP	Global Atmospheric Research Program
GATE	GARP Atmospheric Tropical Experiment
GDAS	Global Data Assimilation System
GFDL	Geophysical Fluid Dynamics Laboratory
GISST	Global Ice and Sea Surface Temperature dataset
GLA	Goddard Laboratory for Atmospheres
GMS	Geostationary Meteorological Satellite
GrADS	Grid Analysis and Display System
GRIB	gridded binary format
GTS	Global Telecommunications System
HIRS	High-Resolution Infrared Sounder
IPCC	Intergovernmental Panel on Climate Change
JMA	Japan Meteorological Agency
KMA	Korean Meteorological Agency
MEDS	Marine Environmental Data Service (Canada)
MSU	microwave sounding unit
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCDC	National Climate Data Center
NCEP	National Centers for Environmental Modeling (formerly NMC)
NESDIS	National Environmental Satellite, Data and Information Service
NH	Northern Hemisphere
NMC	National Meteorological Center (now NCEP)
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
NWS	National Weather Service
OA	optimal averaging
OGP	Office of Global Programs
OI	optimal interpolation
OLR	outgoing longwave radiation
OIQC	OI-based Quality Control
QBO	quasi-biennial oscillation
QC	quality control
SiB	Simple Biosphere Model
SH	Southern Hemisphere
SIRS	Satellite Infrared Spectrometer
SSI	spectral statistical interpolation (also known as a 3D VAR scheme)
SSM/I	Special Sounding Microwave/Imager
SMMR	Scanning Multichannel Microwave Radiometer
SST	sea surface temperature
SSU	stratospheric sounding unit

T62	Triangular 62-waves truncation
TIROS	Television Infrared Observation Satellite
TOVS	TIROS-N Operational Vertical Sounder
TWERLE	Tropospheric Wind Earth Radio Location Experiment
UKMO	United Kingdom Meteorological Office
USAF	United States Air Force
VTPR	Vertical Temperature and Pressure Radiometer
WMO	World Meteorological Organization
XBT	bathythermograph

Appendix E: Content of the NCEP/NCAR Reanalysis Climatology AMS CD-ROM

The enclosed CD-ROM, the first ever included with the *Bulletin of the American Meteorological Society,* includes four types of files: climatologies (13-year average monthly fields), monthly fields (for each of the 13 years), selected daily fields for 1993, and selected observed fields. All the fields have been interpolated to a uniform latitude–longitude 2.5° resolution grid (144 by 73 grid points). The 17 pressure levels are 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 hPa for the climatology and monthly mean fields, and a subset of five levels (850, 700, 500, 200, and 30 hPa) for daily values. There are other single-level fields (e.g., precipitation); isentropic potential vorticity (IPV) on 11 isentropic levels (650, 550, 450, 400, 350, 330, 315, 300, 290, 280, and 270 K) for the monthly fields; and three selected levels (450, 330, and 315 K) for the daily fields.

Because of the horizontal and vertical interpolation, it is recommended that these fields not be used for budget studies, which generally require access to the original data. Z, U, V, T, and MSLP are of type A (analysis variable is strongly influenced by observed data); W, RH, Q, PWAT, U10, V10, T2M, and IPV can be considered of type B (although there are observational data that directly affects the value of the variable, the model also has a very strong influence on the analysis value); most other variables are of type C (indicating that there are no observations directly affecting the variable, so that it is derived solely from the model fields forced by the data assimilation to remain close to the atmosphere.

The following fields are included in the monthly and climatological filedirectories. The * indicates they are also included in the daily (1993) directory file. With the exception of the first seven, these are single-level fields.

Z	Geopotential height (gpm)*
U	u wind (m s^{-1})*
V	v wind (m s^{-1})*
T	Temperature (K)*
W	Pressure vertical velocity (Pa s^{-1})* (500 hPa only)
RH	Relative humidity (%) (below 300 hPa)
Q	Specific humidity (kg kg^{-1}) at 850, 700, and 500 hPa, for daily fields only
PWAT	Precipitable water (kg m^{-2})*
MSLP	Pressure reduced to MSL (Pa)*
CPRATE	Convective precipitation rate (kg m^{-2} s^{-1})*
CSDLF	Clear sky downward longwave flux (W m^{-2})
CSDSF	Clear sky downward shortwave flux (W m^{-2})
CSULF	Clear sky upward longwave flux (W m^{-2})
CSUSFTOA	Clear sky upward shortwave flux at top of atmosphere (W m^{-2})
CSUSFSFC	Clear sky upward shortwave flux at the surface (W m^{-2})
DLWRFSFC	Downward longwave radiation flux at the surface (W m^{-2})
DSWRFTOA	Downward shortwave radiation flux at top of atmosphere (W m^{-2})*
DSWRFSFC	Downward shortwave radiation flux at the surface (W m^{-2})*
ICEC	Ice concentration (ice=1; no ice=0) (1/0)
LHTFL	Latent heat flux (W m^{-2})*
PRATE	Total precipitation rate (kg m^{-2} s^{-1})
RUNOFF	Runoff (kg m^{-2})*
SFCR	Surface roughness (m)
SHTFL	Sensible heat flux (W m^{-2})*
SOILW10	Volumetric soil moisture content 10-m layer (fraction)*
SOILW200	Volumetric soil moisture content at 200-m layer (fraction)*
Q2M	Specific humidity at 2 m above ground (kg kg^{-1})*
HCLDCOV	High-cloud cover (%)
MCLDCOV	Middle-cloud cover (%)
LCLDCOV	Low-cloud cover (%)
TSFC	Skin temperature (K)*
T2M	Temperature at 2 m above ground (K)*
UGWD	Zonal gravity wave stress (N m^{-2})*

UFLX	Zonal component of momentum flux (N m^{-2})*
U10M	u wind at 10 m above ground (m s^{-1})*
ULWRFTOA	Upward longwave radiation flux at top of the atmosphere OLR (W m^{-2})*
ULWRFSFC	Upward longwave radiation flux at the surface (W m^{-2})*
USWRFTOA	Upward shortwave radiation flux at top of the atmosphere (W m^{-2})*
USWRFSFC	Upward shortwave radiation flux at the surface (W m^{-2})
VGWD	Meridional gravity wave stress (N m^{-2})*
VFLX	Meridional component of momentum flux (N m^{-2})*
V10M	v wind at 10 m above ground (m s^{-1})*

The following components of the heat and moisture budget are only available for the 13-year climatology (17 pressure levels):

LRGHR	Large-scale condensation heating rate (K s^{-1})
CNVHR	Deep convective heating rate (K s^{-1})
SHAHR	Shallow convective heating rate (K s^{-1})
VDFHR	Vertical diffusion heating rate (K s^{-1})
SWHR	Shortwave radiative heating rate (K s^{-1})
LWHR	Longwave radiative heating rate (K s^{-1})

The isentropic potential vorticity is available in the monthly mean (11 levels) and 1993 daily (3 levels):

IPV	Isentropic potential vorticity (m^2 s^{-1} kg).

Fixed fields (type **D**) are included in a separate file.

OROG	Orography (m)
MASK	Land–sea mask (1 for land, 0 for sea).

The observed fields included are

OBS93OLR	Daily values of outgoing longwave radiation for 1993 (W m^{-2})
OBSMNOLR	Monthly means outgoing longwave radiation (W m^{-2})
NCEPRAIN	Xie–Arkin estimated rainfall rates (mm s^{-1})
MRGDRAIN	Schemm estimated rainfall rates (mm s^{-1}).

Measurements of OLR are obtained from the Advanced Very High Resolution Radiometer (AVHRR) aboard the NOAA polar orbiting spacecraft (Gruber and Krueger 1984). The data units are W m^{-2} and each value represents the areal average OLR flux for a 2.5° × 2.5° "box." The observations of OLR during the 1979–94 time period that are included on this CD-ROM are exclusively from the "afternoon" satellite; that is, one which observes at the Equator near 0230/1430 LST. It should be noted that considerable observing time drift occurs during the lifetime of the afternoon polar orbiting satellites, and observing times can be up to 5 hours later toward the end of a satellites lifetime compared to the initial launch observing time.

The Xie–Arkin precipitation analysis (Xie and Arkin 1996) is derived in a two-stage process from monthly rain gauge observations and several estimates based on satellite data. First, the satellite estimates are combined using a weighted average where the weights are proportional to the estimated errors of the various estimates. This weighted average is then merged with an analysis of gauge observations over land and with observations from atoll gauges over the ocean. In general, gauge values are used wherever available.

The merged precipitation dataset for 1979–1992, prepared by J.-K. E. Schemm, was generated by combining observed monthly total precipitation data from the world surface station climatology (Spengler and Jenne 1990) from NCAR and estimated oceanic precipitation from the MSU measurements (Spencer 1993). The station data were interpolated to a resolution of 2.5° longitude–latitude by averaging station values within a 200-km radius with weights proportional to the inverse of square distance (Schemm et al. 1992). An attempt was made to control the quality of the dataset by removing station data reporting total precipitation amounts over 1000 mm. The MSU estimates were screened for sea ice contamination by removing data with monthly totals greater than 900 mm in regions poleward of 50° latitude.

The Global Precipitation Climatology Project, which is administered by the Global Energy and Water cycle Experiment, has produced a monthly mean 2.5° gridded precipitation dataset for the period July 1987–December 1994 (December 1987 is missing). This dataset has been produced by blending gauge and infrared and microwave satellite estimates of precipitation. While the instantaneous microwave-based precipitation estimates are more accurate than IR-based estimates, the microwave estimates suffer from reduced temporal sampling (twice daily) relative to the

IR (eight times daily) due to the polar orbit of the spacecraft that house the SSM/I instruments (most of the IR data are from geostationary satellites). Thus, an adjustment procedure has been developed that attempts to meld the strengths of these two estimates—that is, increased accuracy from the microwave combined with better temporal sampling from the IR.

The adjustment procedure is an adaptation of earlier work by Huffman et al. (1995), and consists of steps that first remove the biases in infrared estimates by adjusting to coincident microwave estimates of precipitation. The microwave estimates are obtained from the SSM/I instrument aboard the Defense Meteorological Satellite Program series of satellites and utilize a scattering model for estimates over land and an emission model for over ocean estimates. The final analysis step adjusts the merged satellite data to the gauge observations and combines them using weights that depend on the estimated local error of each field. The gauge data are analyses from the Global Precipitation Climatology Centre and reflect approximately 6700 gauges that have been carefully quality controlled.

This is a new dataset and we request that users provide comments about it to Arnold Gruber, manager of the GPCP, at agruber@orbit.nesdis.noaa.gov. For more information about the Global Precipitation Climatology Project, see the GPCP home page on the World Wide Web: http://orbit-net.nesdis.noaa.gov/gpcp/.

References

Campana K. A., Y.-T. Hou, K. E. Mitchell, S.-K. Yang, and R. Cullather, 1994: Improved diagnostic cloud parameterization in NMC's global model. Preprints, *10th Conf. on Numerical Weather Prediction,* Portland OR, Amer. Meteor. Soc., 324–325.

Chelliah, M., 1994: An assessment of the monthly mean atmospheric fields from the NMC/NCAR Reanalysis Project. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, U.S. Dept. of Commerce, 230–233.

Collins, W. G., and L. S. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Wea. Rev.,* **118,** 2754–2767.

——, and ——, 1992: Complex quality control of rawinsonde heights and temperatures (CQCHT) at the National Meteorological Center. NMC Office Note 390, 30 pp. [Available from NOAA/NCEP, 5200 Auth Rd., Washington, DC, 20233.]

Derber, J. D., and A. Rosatti, 1989. A global oceanic data assimilation system. *J. Phys. Oceanogr.,* **19,** 1333–1347.

——, D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Wea. Forecasting,* **6,** 538–547.

——, D. Parrish, W.-S. Wu, Z.-X. Pu, and S. Rizvi, 1994: Improvements to the operational SSI global analysis system. Preprints, *10th Conf. on Numerical Weather Prediction,* Portland, OR, Amer. Meteor. Soc., 149–150.

Dorman, J. L., and P. J. Sellers, 1989: A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the Simple Biosphere Model (SB). *J. Appl. Meteor.,* **28,** 833–855.

Finger, F. G., M. E. Gelman, J. D. Wild, M. L. Chanin, A. Hauchecorne, and A. J. Miller, 1993: Evaluation of NMC upper-stratospheric temperature analyses using rocketsondes and lidar data. *Bull. Amer. Meteor. Soc.,* **74,** 789–799.

Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.,* **116,** 1138–1156.

——, 1993: Optimal averaging of meteorological fields. NMC Office Note 397, 45 pp. [Available from NOAA/NCEP, 5200 Auth Rd., Washington, DC 20233.]

Goodberlet, M. A., C. T. Swift, and J. C. Wilkerson, 1989: Remote sensing of ocean surface winds with the special sensor microwave/imager. *J. Geophys. Res.,* **94,** 14 547–14 555.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.,* **121,** 764–787.

Gruber, A., and A. F. Krueger, 1984: The status of the NOAA outgoing longwave radiation data set. *Bull. Amer. Meteor. Soc.,* **65,** 958-962.

Huffman, G. J., R. F. Adler, B. Rudolf, U. Schneider and P. R. Keehn, 1995: Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis and NWP model precipitation estimates. *J. Climate,* **8,** 1284–1295.

IPCC, 1990: Climate change: The IPCC scientific assessment. WMO/UNEP International Panel on Climate Change, Doc. 9, 50 pp. [Available from WMO, Case Postale No. 2300, CH-1211 Geneva 2, Switzerland.]

Jaeger, L., 1976: *Monatskarten des Niederschlags fur die ganze Erde, Berichte ses Deutscher Wetterdienstes,* Nr. 139 (Band 18). Offenbach, A. M., 33 pp. and plates.

Janowiak, J., 1994: An evaluation of NMC/NCAR reanalysis tropical rainfall during 1986–88. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, U.S. Dept. of Commerce, 238–241.

——, and P. A. Arkin, 1991: Rainfall variations in the tropical Pacific during 1986–1989, as estimated from observations of cloud-top temperature. *J. Geophys. Res.,* **96**(Suppl.), 3359–3373.

Ji, M., A. Kumar, and A. Leetmaa, 1994: A multiseason climate forecast system at the National Meteorological Center. *Bull. Amer. Meteor. Soc.,* **75,** 557–569.

Kalnay E., and R. Jenne, 1991: Summary of the NMC/NCAR reanalysis workshop of April 1991. *Bull. Amer. Meteor. Soc.,* **72,** 1897–1904.

——, and Coauthors, 1993: The NMC/NCAR CDAS/Reanalysis Project. NMC Office Note 401, 42 pp. [Available from NOAA/NCEP, 5200 Auth Rd., Washington, DC 20233.]

Kanamitsu, M., 1989: Description of the NMC global data assimilation and forecast system. *Wea. Forecasting,* **4,** 334–342.

——, and Coauthors, 1991: Recent changes implemented into the global forecast system at NMC. *Wea. Forecasting,* **6,** 425–435.

Kistler, R., S. Saha, and J. Woollen, 1994: The NMC/NCAR reanalysis monitoring system. *Proc. 19th Annual Climate Di-*

agnostics Workshop, College Park, MD, U.S. Dept. of Commerce, 226–229.

Krasnopolsky, V. M., L. C. Breaker, and W. H. Gemmill, 1995. A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the SSM/I. *J. Geophys. Res.,* **100,** 11 033–11 045.

Legates, D. R., 1993: Global observations, analyses and simulation of precipitation. World Meteorological Organization, WMO/TD S44, 23–29.

——, and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.,* **10,** 111–127.

Levitus, S. 1982. Climatological atlas of the world ocean. NOAA Prof. Paper 13, U.S. Government Printing Office, 173 pp, plus 17 fiche.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **109,** 701–721.

Mahrt, L., and H.-L. Pan, 1984: A two-layer model of soil hydrology. *Bound.-Layer Meteor.,* **29,** 1–20.

Matthews, E., 1985: Atlas of archived vegetation, land-use and seasonal albedo data sets. NASA Tech. Memo. 86199, 53 pp. [Available from Goddard Space Flight Center, Greenbelt, MD 20771.]

Mo, K. C., X. L. Wang, R. Kistler, M. Kanamitsu, and E. Kalnay, 1995: Impact of satellite data on the CDAS-reanalysis system. *Mon. Wea. Rev.,* **123,** 124–139.

Morel, P., 1994: Scientific issues underlying the global energy and water cycle. Abstracts, *European Conf. on the Global Energy and Water Cycle,* London, The Royal Meteorological Society, p. 23.

NOAA/NMC Development Division, 1988: Documentation of the NMC global model. 244 pp. [Available from NOAA/NCEP Environmental Modeling Center, 5200 Auth Rd., Washington, DC 20233.]

Pan, H.-L., and L. Mahrt, 1987: Interaction between soil hydrology and boundary-layer development. *Bound.-Layer Meteor.,* **38,** 185–220.

——, and Wan-Shu Wu, 1994: Implementing a mass flux convective parameterization package for the NMC medium-range forecast model. Preprints, *10th Conf. on Numerical Weather Prediction,* Portland, OR, Amer. Meteor. Soc., 96–98.

Parker, D. E., C. K. Folland, M. N. Ward, K. Maskell, A. Bevan, and M. Jackson, 1993: Ocean–atmosphere climatic fluctuations on interannual to century time scales. Rep. 69, Hadley Centre for Climate Prediction and Research, Bracknell, England, 69 pp.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.

——, ——, J. Purser, W. Wu, and Z. Pu, 1996: The NMC Global Analysis System: Recent improvements and future plans. *J. Meteor. Soc. Japan,* in press.

Ramanathan, V., B. R. Bartstrom, and E. F. Harrison, 1989: Climate and Earth's radiation budget. *Phys. Today,* **42,** 22–32.

Rasmusson, E. M., and T. H. Carpenter, 1983: The relationship between eastern equatorial Pacific sea surface temperatures and rainfall over India and Sri Lanka. *Mon. Wea. Rev.,* **111,** 517–528.

Reed, R., and D. G. Rogers, 1962: Circulation of the tropical atmosphere in the years 1954–1960. *J. Atmos. Sci.,* **19,** 127–135.

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate,* **1,** 75–86.

——, 1993: Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperatures. *J. Climate,* **6,** 768–774.

——, and D. C. Marsico, 1993: An improved real-time global sea surface temperature analysis. *J. Climate,* **6,** 768–774.

——, and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate,* **7,** 929–948.

Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.,* **115,** 1606–1626.

——, and ——, 1989: Precipitation patterns associated with the high index phase of the Southern Oscillation. *J. Climate,* **2,** 268–284.

Saha, S., and M. Chelliah, 1993: Automatic monitoring system for the atmospheric reanalysis project at NMC. *Proc. 18th Annual Climate Diagnostics Workshop,* Boulder, CO, U.S. Dept. of Commerce, 330–333.

——, M. Kanamitsu, M. Chelliah, and C. Ropelewski, 1994: Surface precipitation over the U.S. from the NMC/NCAR reanalysis. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, U.S. Dept. of Commerce, 234–237.

Salstein, D., 1993: Temporal and spatial resolution impact on some climate analysis diagnostics. *Proc. 17th Annual Climate Diagnostics Workshop,* Norman, OK, U.S. Dept. of Commerce, 354–359.

Schemm, J., S. Schubert, J. Terry, and S. Bloom, 1992: Estimates of monthly mean soil moisture for 1979–1989. NASA Tech. Memo. 10451, NASA/Goddard Space Flight Center, Geenbelt, MD.

Schubert, S., R. Rood, and J. Pfaendtner, 1993: An assimilated dataset for earth sciences applications. *Bull. Amer. Meteor. Soc.,* **74,** 2331–2342.

Shukla, J., and D. A. Paolino, 1983: The Southern Oscillation and long-range forecasting of the summer monsoon rainfall over India. *Mon. Wea. Rev.,* **111,** 1830–1837.

Smith, T. M., 1994: Preliminary analysis of the annual cycle of vertical heat flux into the Pacific from the NMC/NCAR reanalysis. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, U.S. Dept. of Commerce, 242–245.

Spencer, R. W., 1993: Global oceanic precipitation from the MSU during 1979–1991 and comparisons to other climatologies. *J. Climate,* **6,** 1301–1326.

White, G., 1994: Diagnostics of atmospheric forcing from reanalysis. *Proc. 19th Annual Climate Diagnostics Workshop,* College Park, MD, U.S. Dept. of Commerce, 246–249.

Woollen, J. S., 1991: New NMC operational OI quality control. Preprints, *Ninth Conf. on Numerical Weather Prediction,* Denver, CO, Amer. Meteor. Soc., 24–27.

——, E. Kalnay, L. Gandin, W. Collins, S. Saha, R. Kistler, M. Kanamitsu, and M. Chelliah, 1994: Quality control in the reanalysis system. Preprints, *10th Conf. on Numerical Weather Prediction,* Portland, OR, Amer. Meteor. Soc., 13–14.

Xie, P. P., and P. A. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate,* in press.