The NCAR Status Documents for Reanalysis

Roy Jenne
26 Jan 2001

This bundle of papers has over 25 status documents from NCAR for reanalysis. They document the recent work on reanalysis and the status of all of the datasets as the reanalysis project was progressing. They were written periodically as the NCEP/NCAR 50-year project was being worked on (during 1992 – 98). Production of the NCEP/NCAR reanalyses started June 1994, and 50 years (1948 – 97) was completed July 1998. Then it is updated each month. In Jan 2001, the years 1948 – 2000 (53 years) have been done.

There is information in these documents that does not exist in any other place. A separate bundle of papers has similar information for the more recent period (1998 – 2000)

If you need an overview of the whole reanalysis program, the first documents to read are:

1. The NCEP/NCAR big project: A description of the reanalysis project is in the March 1996 issue of the Bulletin of the AMS.

2. A 20-page paper that summarizes information about the observations (dated Aug 2000). This is an on-line document (and paper) at NCAR.

Total pages in document bundle:
1. Introduction:  3 pages
2. Part 1:  98 pages
3. Part 2:  104 pages
Total:  205 pages
Some Status Documents on Reanalysis

Roy Jenne
19 July 2000
Rev 2 Feb 2001

2. Data Support for Reanalysis Projects (18 Nov 1997) 1 p
3. Status of Reanalysis Data (1 Apr 1993) 6 p
4. The Preparation of Data for Reanalysis (14 Jul 1993) 3 p
5. Some Reanalysis Data Issues (RJ, 11 Nov 1993) 6 p
6. Reanalysis Data (RJ, 26 Jan 1994) 8 p
   - Mostly about the NMC/NCAR project to do 40 years of analyses.
   - Variables are listed.
   - Some information about ECMWF is given.
8. Considerations for Reanalysis (7 Dec 1994) 4 p
10. NMC/NCAR Reanalysis Information (Jan 1995) 2 p
11. Data Status from NMC/NCAR Reanalysis (27 Apr 1995) 9 p
   - This has 5 slides for a meeting.

Begin Part 2

    - This status doc has lots of information
    - In Oct 1995, 12 years are done (NMC/NCAR project)
    - There is information about the obs. Atch 5 is called “Status of Sending Data to NMC”
      (28 Aug 1995).
    - There is some information about reanalysis projects at ECMWF, Goddard and COLA.
    - Some information about observations
    - Some information about output data
15. Reanalysis Update, Dec 13 (13 Dec 1995) 2 p
17. The NCEP/NCAR Project To Reanalyze the Atmosphere (27 Feb 1996) 5 p
18. What is the Status of Reanalysis (27 Mar 1996) 4 p
   - 17.0 years (1979 – 95) were done 11 Feb 1996.
20. Reanalysis (Talk with Kistler) (RJ, 17 June 1996) 3 p
   - About the bad Southern Hemisphere bogus; about SST and sea ice problems.
22. Fixing Old Raobs (14 Feb 1997) 2 p
   - Some data problems we are finding.
25. Data for Reanalysis (13 Mar 1997) 5 p
   - This is 70% different from the 4 Mar 1997 text.
26. NCEP/NCAR Reanalysis (30 May 1997) 2 p
   - From talk with Kistler.
   - Problem of some raob data not used.
28. Comments About Reanalysis Data (9 Nov 1998) 5 p
   - NCEP B-3, loc of ships, station numbers, etc.
   - An important summary
Time Line of Production of NCEP/NCAR Reanalysis

Roy Jenne
19 July 2000

- Feb 1991: The new work started at NCAR (prepare observations) and at NCEP (improve models, process observations from NCAR, and prepare automated production systems).

- June 1994: Production started at NCEP.

- July 1998: The production of 50-years of global analyses (each 6-hours) was completed.

- The resolution: Analyses were done at 28 levels, at a horizontal resolution of 208 km (T62, L28).

- The Time Line of Production
The following page is a time line of the production of reanalysis. For example, it shows that about 2 years of production were done during June – July 1994. Then tests shows that a spike in surface temperature was caused in summer by the way that vegetation (wheat) and soils were handled. Methods were improved and production was restarted. The SSMI surface ocean winds started in mid-1987. These had a bias because of a misunderstanding about the format, etc. So this problem had to be coped with (during the period 2 Dec 94 – 25 Apr 95).
Production of NCEP/NCAR Reanalysis

- **23 July 1998**: 50 years done (1948-97)
- **October 1997**: 40 years done (1957-96)
- **March 1997**: 29 years done
- **February 1996**: 17 years done
- **December 1994**: 5 years done

---

Roy Jenne  
Oct 1997
Chapter V
SCHEDULED TIMES OF OBSERVATIONS

During the period January 1, 1946, through March 31, 1948, the scheduled times of observations were 0400 and 1600 GMT, with delayed observations to be made within the succeeding six hours. On August 8, 1946, stations were granted permission to take only the 1600 GMT observation until four observers were experienced, but the length of time for this was not expected to exceed three weeks. Permission was granted on October 1, 1946, for stations requiring additional time to meet transmission schedules to release one and one-half hours before the scheduled times of release. However, this was not done at all stations.

Scheduled times of release were changed to 0300 and 1500 GMT for two observations per day, and 0300, 0900, 1500, and 2100 GMT for four a day, beginning with the first observation after 0001 GMT, April 1, 1948. Delayed observations were to be made within the succeeding three hours when two or more observations per day were scheduled and within the succeeding six hours when only one per day was scheduled.

By agreement of the World Meteorological Organization, times of upper air observations were changed to 0000, 0600, 1200, and 1800 GMT, where four observations per day are taken, and 0000 and 1200 GMT where two observations per day are taken. It was further provided that releases could be made as much as 60 minutes before scheduled time. Observations started after scheduled times were to be considered as delayed observations and could be taken as late as 60 minutes after schedule. This was effective April 1, 1957, in all regions except the United States, Canada, and the Pacific Ocean, and on June 1, 1957, in those areas.

Page 23

Weather Bureau

KEY TO METEOROLOGICAL RECORDS DOCUMENTATION NO. 5.21

HISTORY
AND
CATALOGUE OF UPPER AIR DATA
FOR
THE PERIOD 1946-1960

Washington DC : 1964
Contact NCDC Asheville

Present practice - for many years the balloons are usually released one hour before the standard time (thus 23.6 GMT and 11.8). In Russia they are released 30 min before the standard time.

-Rudy Jannke
Sep 2000
Data Support for Reanalysis Projects

1. The NCEP/NCAR reanalysis project.
   - Our work started February 1991, based on much prior work.
   - Where does production stand?
     ➢ Production started June 1994
     ➢ February 1996: finished 17 years (1979-95)
     ➢ We still plan to do 1948-1957.

2. Data preparation has been a messy problem.

   All the data decks have some problems. Some have been very messy (data for wrong times were merged together; wrong identification of stations; data had been assigned to wrong levels in raobs, etc.). This work is based on many good past data efforts around the world. Data from about 20 countries and especially USAF and NCDC in Asheville.

   We have used 24-person years on this project; we have had to shift our priorities and resources to get this done.

3. More work to do on reanalysis data.

   We are pleased with the amount of data and the data quality that we have been able to deliver to reanalysis. However, we have a list of things that should still be done to prepare for the future. For example, there has been no way (yet) to obtain raobs for South Africa for 1950-1957 and 1960. The rawinsonde observations were taken.

4. The new ECMWF reanalysis project.

   ECMWF has plans to do a 40-year reanalysis (1957-on). In September 1997, we completed agreements with them. We will give them the necessary observations to do it. They will give us outputs from the present 15-year ERA-15, and the new ERA-40, either free or at a reduced price. We will do a lot of work. We will save $1 million or 2 million compared with their normal prices. It will help give the world two long reanalysis projects.

5. The observations (50-years) are like a pot of gold.

   The observations will serve the needs of future reanalysis efforts. Also they will help much other research.

   - The observations include rawinsondes, pibals, aircraft reports, surface land, surface marine (COADS), satellite sounders (Apr 1969-on), satellite cloud winds, constant level balloons, etc.
   - More data gathering and clean-up work is still needed, but the world now has a solid basis for much research.


   Many people want to understand more about climate variability. The reanalysis projects are critical to these needs. In January 1996, the NCEP/NCAR project was extended to include the earlier years 1948-1956. In summer 1996, Lennart Bengtsson gave a strong pitch for long reanalysis from two centers. By November 1997, several people (and a panel) are pushing hard for earlier reanalyses if at all possible, especially for 1930-1947.
Status of Reanalysis Data

1. COADS

All the data for 1980 through 1992 is sorted and ready for dupe elimination. 10.195 Gbytes of data were input and 8.093 Gbytes were output for 1980-1992, and the rest is earlier data. Dupe elimination will get rid of half or more of the data.

- Data for the 1980s will be ready about May 15, 1993
- Data for the 1970s will be ready about Nov 1, 1993
- Data for the earlier part of the whole period (1947-1992) will be ready about Feb 1994
- We should make special checks on data for 1979. How does COADS compare with FGGE?

2. Rawinsondes (NMC data for 1962-1992)

Gregg Walters has finished looking at inventories for 1973-1992, and has 25 pages of notes of cases that possibly have trouble with headers, dates, etc.

In late Feb 1993, Dennis Joseph prepared the rawinsonde and wind files for Jan and Feb 1985. He still has a few more checks to make. He has plotted data using GrADS, etc. About Apr 10 we could send a lot of data for 1985-on. We prefer to first send 2 months of data and get feedback before we continue.

Will Spangler is about to tackle the date and time problems that exist in some of the NMC data during 1962-1972.

3. Rawinsonde Data for 1948-on (USAF, TD54)

NCAR has some of this data converted from earlier work. We sent the whole TD54 to GFDL, and they have done a lot of work to get over the first set of problems. By about Jul 1993, they will be able to make the next necessary series of checks.

4. Station Elevation for Raobs

We (Joey Comeaux) have calculated surface elevation from all of the NMC raobs during 1973-1992. The plots show that the elevation data on the tapes is often wrong. We will gradually derive correct elevation data, but this cannot be done before we send the basic tapes to NMC. We need to discuss this point, because correct information will need to be added later.

5. Surface Land Synop

a. We have 3-hour NMC data for 1976-present, but have not started to prepare it yet.
b. We (Ilana Stern) are working on the land synop data for Jan 1967 - Dec 1980 (from USAF decode). The month of Nov 1970 was missing, but it is now filled in. The data does not have station location data now, but we hope that a tape coming from AF/ETAC will have the needed data.

c. Synop data for 1945-1971 - This TD-13 data from the USAF arrived at NCAR in Aug 1990 (on 405 tapes). We had to key-enter station information and understand many formats. About 80% of the work has been done. We have to check for nonstandard codes and build routines to unpack the data in a uniform way.

6. Aircraft Data

We (Bob Dattore) have almost finished adding several new datasets. Now the GATE research data is nearly done. Next we will prepare GATE commercial aircraft data. We are still waiting for the data from New Zealand.

• I also want to include TWERLE balloon data (Jul 1975 - Aug 1976, and some data later). These data are enough to establish quite a good reference level in the S. Hemisphere at about 150 mb. Each report is like one mandatory level of a rawinsonde report, located near 150 mb.

7. TOVS 2.5° Data

NCAR will soon obtain a copy of TOVS 2.5° data from SDSD. The data at NMC was obtained from NESDIS/research. NCAR will prepare an inventory and compare it with the NMC inventory.

We know that there are problems with week-long gaps in the 2.5° data. The basic TOVS data has very few gaps. All of us need to invent a way to provide a complete record for reanalysis.

8. Tropical Storm Data

Charlie Neumann called on Mar 30. Some of the data from Guam was not good. It will probably take 6 weeks to obtain new data and process it. In the various ocean basins, the data will then be updated either through Dec 1992 or mid-1992.

- End -
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>'46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'57</td>
<td>U.S. Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'55</td>
<td>West Coast of S. America</td>
<td></td>
<td></td>
<td>'87</td>
<td></td>
</tr>
<tr>
<td>'43</td>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
<td>'87</td>
</tr>
<tr>
<td>'55</td>
<td>New Zealand</td>
<td></td>
<td>'75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>French Polynesia</td>
<td></td>
<td></td>
<td>'73</td>
<td></td>
</tr>
<tr>
<td>'51</td>
<td>French Africa</td>
<td></td>
<td>'65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Africa</td>
<td></td>
<td></td>
<td>'89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UK overseas</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'53</td>
<td>Ascension Island</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'57</td>
<td>Singapore</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>India</td>
<td></td>
<td></td>
<td></td>
<td>'78</td>
</tr>
</tbody>
</table>

Need: Argentian (coming)
Brazil
Others

Roy Jenne
Mar 1993
Aircraft Data

Aircraft and balloon data at NCAR. We also have aircraft data from the Navy. But data from NZ is not here yet.


1947  AF Recon  1959

1976 USAF 1985

1979

GASP

Gate
Jun-Aug 1974

FGGE
1979

TWERLE
Jun-Aug '75 '76

Sadler

Mar 1962

NMC Aircraft

Dec 1971 Australia 1989

NZ

95 90%

10%

0

5%

95

60

0

0

Roy Jenne
Oct 1992
ELEVATION PLOT FOR STATION 22113 for 12z

MEDIAD AND REPORTED ELEVATION PLOT FOR STATION 22113 for 12z

Marinansk, Russia
69.0 N 33.1 E

* This shows that the station elevation really did change near day 5600.
The Preparation of Data for Reanalysis

Surface and upper-air observations are being prepared for use in the planned reanalysis. The plan is to use the data available for the original operational NMC analyses (this data exists from Mar 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. The component datasets are summarized in more detail below, and selected texts with more information are listed.

COADS Surface Marine Data

From about 1982-on there has been a large joint project involving NCAR, ERL (Boulder) and NCDC to prepare the best set of global ocean surface marine data that is possible. Canada and other countries have also helped. The Polar Science Center (University of Washington) has provided Arctic ice buoy data. The COADS dataset includes ships, fixed buoys, drifting buoys, pack ice buoys, near-surface data from ocean station reports (XBTs, etc.), and some other data. NCDC key-enters data from U.S. ship logs; data also comes in from many countries in delayed time. After a wait of 5 years, COADS has about twice the number of ship data, compared to real-time data.

The data for 1854-1979 was released in 1983. A major 1980-92 update was completed in June 1993, after an intensive phase of 2 years. A copy of the recent data is being put into synoptic sort and sent to NMC (July 1993). Reynolds will then prepare the enhanced weekly 1° analyses of sea surface temperatures (1981-92) that are needed for reanalyses.

Additional data are now available for the older COADS period (1854-1979) and are being assembled. A revised COADS set for 1947-79 will be ready to send to NMC about April 1994 to use for reanalysis.

Global Rawinsonde Data

NCAR has tapes of the NMC GTS data with upper-air observations from Mar 1962-on. These are all original tapes from NMC, so no earlier conversions were made (only migration to new media). These data will be used for reanalysis. We plan to provide both the GTS data (which also has pibal and aircraft) and also raobs from national archives in various countries. A few days of GTS data have the wrong dates, which will be corrected. NCAR has raobs directly from some countries such as South Africa, Australia, Canada, and from the U.S. (NCDC). The 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 helping with some of the checking between different sources of the same data.

NCAR, NCDC, Russia, Europe, and others (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. The results will be available for later reanalyses, not the first one. The collaboration should mainly be to prepare the component subsets of data correctly. Different methods will probably be used to accomplish the larger data merges of subsets.
Aircraft Data

Aircraft data are available from real-time GTS sources from NMC, from Mar 1962-on. The data for 1962-72 was one of the types of data that NCAR received on 2000 tapes from NMC about 1973. Data from several other sources is being prepared, including USAF data (1947-59) and tropical aircraft data from Sadler. Both research and commercial aircraft for the tropical Atlantic GATE experiment (1974) will be available. We want the GATE analyses to be especially good, because there was also weather radar data that will help to check cumulus parameterization methods in models. Data from TWERLE constant-pressure balloons for the S. Hemisphere (Jul 1975 - Aug 1976) will be in the dataset. These balloons provide data like one level of a rawinsonde near 150 mb. We may be able to include aircraft data from New Zealand, where only half of the reports ever got onto GTS.

Surface Land Synoptic Data

Data from three main sources will be available by about May 1994 (or earlier) to provide fairly good coverage from 1949-on. These sources are the NMC decode of GTS data from Jul 1976-present; a dataset that R. Davis at NCDC prepared from the USAF GTS data for 1967-80, and a set of world synop data from the USAF (TD13) for earlier years. There are data sources to make further improvements, but they cannot be included in time for the first reanalysis runs.

Satellite Sounder Data

The basic radiances are available for the following:

- SIRS IR sounders
- VTPR IR sounders
- TOVS sounders (IR, Vis, MSU, SSU)
- HIRS data test

<table>
<thead>
<tr>
<th>Product</th>
<th>Available Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIRS IR sounders</td>
<td>Apr 1969 - Apr 1971</td>
</tr>
<tr>
<td>VTPR IR sounders</td>
<td>Nov 1972 - Feb 1979</td>
</tr>
<tr>
<td>TOVS sounders</td>
<td>Nov 1978 - present</td>
</tr>
<tr>
<td>HIRS data test</td>
<td>Aug 1975 - Mar 1976</td>
</tr>
</tbody>
</table>

The first three of these archives are at NCAR. The VTPR archives include basic data and the operational retrieved soundings. For TOVS, there is a separate archive of cloud-cleared radiances and associated retrievals (2.5° resolution). There are some bad gaps in the 2.5° archive, but the archive of TOVS radiances is almost complete.

Satellite Cloud Drift Winds

These wind data from the original NMC tapes will be available for use. Some of the same (but sometimes more complete) data from national archives may also be used.

Timing

The data for 1980-92 should be at NMC by at least Dec 1993, and the first version of the data for 1948-79 should be ready by mid-1994.
Coverage of Data for Reanalysis

A text entitled "Data for Reanalysis; Inventories" (about 115 pages, dated Nov 1992) has various maps and displays that give a feeling for the coverage of 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.

Other Information about Data for Reanalysis

Many reports have been prepared that give more information about the attributes of different datasets and the status of projects to prepare the data. Sea surface temperatures (SST) are needed for reanalysis; sea ice data is needed to prepare SST analyses. Tropical storm locations are needed also. Papers have been prepared that focus on different issues; a selection of these papers is given in Table 1.

<table>
<thead>
<tr>
<th>Text</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea Surface Temperature Data</td>
<td>1 Feb 1993</td>
</tr>
<tr>
<td>Sea Ice Data</td>
<td>2 Apr 1993</td>
</tr>
<tr>
<td>Dataset of Tropical Storm Locations</td>
<td>26 Jan 1993</td>
</tr>
<tr>
<td>TOVS Sounding Data</td>
<td>25 Jan 1993</td>
</tr>
<tr>
<td>Missing Data in the TOVS 2.5° Archive</td>
<td>1 Mar 1993</td>
</tr>
<tr>
<td>Analyses for S. Hemisphere, 1951-on</td>
<td>18 Mar 1993</td>
</tr>
<tr>
<td>NMC Upper Air Data, Inventories 1973-on</td>
<td>12 Mar 1993</td>
</tr>
<tr>
<td>Status of Reanalysis Data</td>
<td>1 Apr 1993</td>
</tr>
</tbody>
</table>
Some Reanalysis Data Issues

These notes summarize and update information about datasets being prepared for use in reanalysis programs. A review meeting for the NMC/NCAR reanalysis program was held at NMC on Oct 20-22, 1993. ECMWF was represented by Per Kallberg. These notes also follow up on discussions at that meeting.

1. An Overview of Data Being Prepared for Reanalysis

Global surface and upper-air observations are needed for reanalysis. Data from real-time telecom sources (GTS) will be used in addition to data from delayed sources. The following main types of data are being prepared, and approximate dates for the data are given:

- Global rawinsondes and upper winds (e.g., pibals), starting 1946 to 1950.
- Aircraft and reconnaissance data start 1947. GTS sources start Mar 1962. Some constant level balloon data are included.
- Satellite sounding data starts Apr 1969.
- Global surface marine (ships and buoys) and pack ice buoys. Develop a new, more complete version of COADS for 1947-on.
- Global land surface synoptic data. Coverage is good for Eurasia and Australia by 1948. GTS sources start mostly in 1967.


The following sections give more information about the status of individual components of the data for reanalysis.

2. Use of Satellite Sounder Data

Data from satellite sounders started Apr 1969, with a gap from May 1971 - Oct 1972.

2.1 TOVS 2.5° Archive (Nov 1978-present)

TOVS has data from three sounders: HIRS (IR and VIS), MSU (microwave) and SSU (stratosphere). The NOAA TOVS archive started in late Oct 1978. For reanalysis, the 2.5° archive of cloud-cleared radiances and associated soundings will be used.

The TOVS 2.5° archive still has lots of gaps. A plan is in place to fill as many gaps as possible by using the best of NMC and NCAR archives. (The origin for the NCAR set was SDSD and NCDC.) SDSD and NCDC-Asheville have copied the archive twice in recent years and each time there are different gaps. But some gaps are real gaps and stay constant. The origin of the NMC set was somewhat different.
In Oct 1993, ECMWF said that they would do cloud clearing and generate soundings to fill the gaps if NCAR provides the raw TOVS radiances. This project is starting (2 Nov 1993).

- NMC will send their archive of 2.5° TOVS data to NCAR.
- NCAR will make a combined 2.5° TOVS archive with the fewest gaps possible.
- NCAR and ECMWF are working together to fill the gaps in the 2.5° TOVS data.
- Both NMC and ECMWF will have the resulting data, which will be an almost complete archive of 2.5° TOVS data for about 1 Nov 1978 to present. The recent cutoff date is still uncertain.
  - We should try for good TOVS data at least through 1992. Could NMC send their data to NCAR at least through Dec 1992?
- In the reanalysis, NMC will use the soundings. ECMWF will probably use cloud-cleared radiances.

2.2 VTPR IR Sounder Data

VTPR IR sounder data overlaps TOVS. NCAR has files of raw data and soundings for Nov 1972 - Feb 1979. We have inventories of the raw data. We will soon make inventories of the soundings to verify that they are as complete as the raw data. The plan is that NMC will use the soundings.

+ Done 12 Nov 93

2.3 SIRS IR Sounders

NCAR has an archive of the raw data from SIRS IR sounders for Apr 1969 - Apr 1971. Jenne should talk with Bill Smith (Wisconsin) to get a scientific assessment of this data. Can we invent a good way to do cloud clearing?

2.4 HIRS Dataset in 1970s

The HIRS dataset for Aug 1975 - Mar 1976 was used for the NASA data system test. It is good data—better than VTPR. It won't be used on this reanalysis because of lack of time. This dataset is not at NCAR.

3. Rawinsonde Data

Two main sources will be used: A basic data stream that mainly has data from the NMC decode of GTS data, and the other is from a lot of other sources (generally not GTS).

- Basic data streams of raobs
  - NMC and upper air data (4 Mar 1962-on)
  - MIT set (May 1958 - Apr 1963)
  - TD54 data (about 1948 to 1970, or some quit earlier)
Data from NMC decode: NCAR has this original NMC data for 4 Mar 1962-present (coverage became global in Jun 1966). These tapes have separate wind observations (pibal or radar) as well as rawinsondes.

- Other rawinsonde data. NCAR has data from a number of countries and several projects. We will soon work to include:
  - GATE dropsondes and ship raobs (summer 1974)
  - Line Islands Experiment data (Mar-Apr 1967)
  - Reconnaissance dropsondes for Arctic (1950-61)
  - Ice Island T3 - check
  - Russian North Pole stations - check

One country that we still hope to obtain data from is Argentina. NCAR will probably combine all of these "other" sources into one data stream for reanalysis.

- Antarctic raobs
  - NCAR has a special inventory of these raobs
  - We recently added two stations from the UK
  - We hope to get the Russian raobs, but this may take 6 to 18 months

4. Surface Marine Data (COADS)

COADS has ship data, drifting buoys, fixed buoys, and some pack ice buoys (for the Arctic). The ice buoys are only partly included now.

- The data for 1980-1992 are ready now (since Jul 1993). This is the best dataset for both NMC and ECMWF to use.
- NCAR, ERL, and NCDC are working on data for 1947-79. NMC will use this.
- For the FGGE year (Dec 1978 - Nov 1979), it is still unclear whether to advise people to use the FGGE data or COADS data. NCAR and ERL are making diagnostic checks to answer this question.

Note: There is still some ice buoy data for the Arctic that NCAR/ERL may be able to add.

5. Aircraft Data

There will be NMC GTS aircraft data from 1972-on.

NCAR has been preparing many other sets of aircraft data:
- These data will all be put into one data stream in synop sort
- The Australian data we still have is in GTS format and needs to be decoded (talk with Cliff Dye at NMC, x8115, about this)
- New Zealand aircraft data. On Nov 3, 1993, I heard from John Kidson from New Zealand. He is trying to speed up the process of getting us the data.
Japan apparently has more aircraft data than does ECMWF. Need to decide about this. ECMWF suggested that NCAR could make a proposal on what to do.

6. Cloud Drift Winds

NCAR has:
- All data on the NMC GTS tapes. Some of this goes back to about 1967, I think.
- Files from SDSD from U.S. GOES satellites (this starts 1974)
- Japan sent their cloud winds from GMS on 28 tapes to NMC (about Aug 1993)
  - NMC will send a copy to NCAR
  - Data is for 6 Apr 1978 through Dec 1991; in 1984, 2 months of data are missing; and there is one other smaller gap

The cloud winds are biased toward speeds that are too low. The bias has improved with time. For the U.S., Europe and Japan, we need to obtain figures that have plots of the bias vs. time.

Kistler said that part (or all) of the wind bias problem for Japan was that fixed heights were being used that were too low (and too warm). Radiance from GMS would still be available to help correct this. Question: "What is the present status of this cloud wind data from Japan?" (Is it the same as the original GTS data?)

7. Surface Land 3-hour Synop

The main big sources of data are:
- NMC for Jul 1976-present
- USAF/NCDC for Jan 1967 - Dec 1980 (Dick Davis set)
- USAF TD13 for 1948-70

NCAR will send the following to NMC:
- NMC data for Jan 1981-present
- Dick Davis set for Jan 1967 - Dec 1980
  - But we still might use NMC in the overlap period (Jul 1976 - Dec 1980)
- TD13 for earlier data
- The sort of the data. All of the data will be sent in synop sort (all world data together for one time). TD13 is now in a time series sort, but NCAR will put it into a synop sort.
- Station library information. Lat-lon and elevation information is not in the Dick Davis set or TD13. NCAR will put it in. I'm still worried about some wrong locations that have always been in the NMC set.

8. Hurricane Location Data

A long global dataset of these hurricane centers will soon become available (from Charles Neumann). There is still some debate on whether or not to use bogus hurricane data, based
on the data for hurricane centers. I think that it would be best to use the centers, if programs can be checked out. If centers are not used, the analyses will still be getting some scraps of observations about the hurricane, and therefore it will be analyzed in a fairly random way, and the location and intensity will bounce around. Also the first guess will be worse. We also learned that the hurricane forecasts are better in high-resolution systems than at low resolution—as one might expect.

- **Update Nov 8, 1993**: Charlie Neumann has the dataset together but he is still doing a few small things to it. This set has mostly 6-hour storm data for many years. Central pressure is given. Storm intensity data is given by max wind estimates and these are usually available: In the Atlantic, these data on intensity are there from the late 1800s. Elsewhere, the data are included from the late 1950s, except the Indian Ocean max winds didn't start until about 1978.

- NMC and ECMWF can start using the data any time. All data can be generally released in Dec 1995; all data except for the Atlantic Ocean can be released now.

9. **Data from Ice Cap Pressure Buoys**

The University of Wisconsin (Stearns) has had ice cap buoys on Antarctica and on Greenland (near the center of the ice cap). Some of this data got on GTS but not all. NCAR has 3-hourly data from the University of Wisconsin. The University of Wisconsin thinks that they now know the elevation of the stations within 10 meters or so.

10. **Australian Bogus Data**

I think that this bogus data from Australia has both SLP and thickness information (to 500 mb?)

- Bogus data on NMC tapes at NCAR will be given back to NMC.
- Per Kallberg (ECMWF) thinks the Australian bogus data started in 1972. I believe that ECMWF got a tape with bogus data, directly from Australia, probably for 1979-on.
- Note that NCAR has analyses for the S. Hemisphere, from Australia, that start Apr 24, 1972.

11. **Variables in Reanalysis Archives**

The archive plans for NMC reanalysis are looking good, but there are still a few quantities to think about:

Dew point at 2m. A moisture archive is planned for the air moisture at the first sigma level. It would be good to have moisture at 2m (we already have temperature). SLP is a popular field that is not yet in the archive.
12. FGGE (Dec 1978 through Nov 1979)

The centers should use the last version of FGGE data (it was ready about 1984, I think). The first version had some troubles.

- But use the TOVS 2.5° data from the present data exercise.
- Use aircraft data from FGGE, but NCAR may need to add some more aircraft data.
- Ship and buoy data: COADS probably has some data not in the FGGE set, but there are some issues that NCAR still needs to sort out.
- Cloud winds: this FGGE set is best source. It has the higher resolution calculations, and it has data from the GOES satellite over India.
- On Nov 9, 1993, Kanamitsu said that ECMWF will provide NMC with FGGE data in BUFR in exchange for the COADS data. Therefore, NCAR will not need to send COADS data to ECMWF.

13. Issues about Data Distribution

- We need to update our information about the volume of data out of reanalysis.
- Jenne has information about where Exabyte technology will be in a year (without compression, 5 Gbytes now fit on a tape and 20 Gbytes will fit in a year).

14. Issue Papers

NCAR has a series of papers with more information about given types of data: hurricane locations, SST, TOVS 2.5°, sea ice, etc.

15. When will Reanalysis Programs Start in Production?

- NMC will start about Apr 1994 (for 1958-92 data). More recent years will be done first.
  - An NMC document dated about Aug 1993 has a lot of information about the project.
- ECMWF will start about Dec 1993 (for 1979-on data)
  - ECMWF has a document (1993) that describes their project.

- End -
Reanalysis Data

This gives the present status of data for reanalysis and helps to answer whether production can start in early April 1994.

1. Grouping of Data

The ideal case would be to have all of the data in 6-hour groups for the upper air, and in 3-hour groups for the surface, all in synoptic sort. The data will all be sent to NMC in synoptic sort, but there are still a few questions about the time grouping.

On the NMC tapes from 1962-on, there is a changing history of how the data have been grouped.

NCAR supplies the surface COADS data for reanalysis in a synoptic sort, so it is easy to pick out 3- or 6-hour periods as desired. The data are all in order by time.

2. Surface marine (ship and buoy data)

We will first discuss the COADS ship and buoy data so that it doesn’t have to be discussed twice in other sections. These data should be ready in ample time to support reanalysis efforts.

   a. NMC has COADS data for 1980-92 from NCAR
      — They will copy it for ECMWF
   b. NCAR, ERL, and NCDC are working on an improved COADS set for 1947-79. This still requires much work by NCAR and ERL. NCAR will put it into synoptic sort.
   c. FGGE data. There are one or two fixed buoys in wild locations in the final FGGE dataset. NCAR is developing more information.
   d. COADS data for 1947-79. This older COADS data will be ready to send to NMC about Aug 1994. It may be delayed until later if NMC doesn’t need it that soon.
   e. COADS data for 1993: NCAR will check on when we could provide data for COADS for 1993 that is better than GTS. We will probably prepare a new set for 1992-93 at this same time, to get more of the data for 1992.

3. Sending data to NMC

NMC wants to start production by 1 Apr 1993. Wait a minute, do we want to start on April fools day? Can data for 1985-92 be at NMC in time to start in early April? There are still a lot of things to do, but NCAR should really start to move data to NMC. Following is summary information about data for 1985-92; more details will follow later in the text.
Raobs

- NCAR should start sending NMC raobs for 1985-on to NMC. Is NMC ready? We will follow with a short list of station lat-longs to correct. There have been some jumps on the NMC tapes.
- There are always a few junk raobs on the NMC data tapes where the station number of a good raob has been garbled to a wrong location. There is a location for these raobs, but it is wrong. We need to discuss how we will tell programs not to use them.
- The supplemental raobs will come somewhat later

Aircraft

- NCAR should start sending NMC aircraft data to NMC about Feb 1, 1994
- Supplemental aircraft data will come a little later. We need to discuss how much of this to include for recent years.

Cloud winds

- NCAR should start sending these NMC data about Feb 15.
- Need to discuss the supplemental data
- NMC has a copy of Japan’s data (6 Apr 1978 thru Dec 1991)

Special cloud winds for hurricanes

The Hurricane Center has been making a special set of cloud wind data for their area of interest, in order to help the analysis of hurricanes. Kistler will try to get this dataset.

TOVS 2.5° data

- We need the tapes from NMC soon. Then we will do the merges. Jenne made an estimate of resulting gaps that will still exist; NMC should already have a copy of this list. ECMWF will help to fill the gaps.
- NMC mailed the TOVS tapes about 25 Jan 1994.
- Bob Kistler says this is about 150 tapes, in the original format, on 3480 cartridges.
- Bob will check on the status of an update for the last 2 years.

Surface 3-hour land

NCAR could send some of this data soon. Start sending lots of data for 1985-on about Mar 1. NCAR is still doing diagnostics to detect and solve location problems (lat-lon).

- We need to discuss the 3- and 6-hour data options for 1979-82.
- No supplemental surface land data is needed.
- NCAR has on-line inventory files with a print line for each day that shows the count of reports each 3 hours. One file has data for Jul 1976 to Dec 1993 from NMC. The other file is from Jan 1967 to Dec 1980 from USAF GTS, via NCDC-Asheville.
Bogus data

On Mar 11, 1986, NMC stopped sending the bogus file to NCAR. However, I think that NMC has the tape of Australian bogus data via ECMWF. True?

- NMC does not want the bogus file from NMC
- Kistler has Australian bogus data for 1978-on, via ECMWF
- NCAR will ask Australia for earlier bogus data

Ice cap data

Jenne needs to check on updates with the University of Wisconsin.

Tropical storm data for many years

This is ready to send. NMC got the file about mid-Jan 1994

New NMC archive

Kistler started saving his own copy of NMC operational data on 1 Jan 1994.

Note about the location of data

Because of the garbling of station numbers and errors in station location lists, there are always a few big errors in the Lat-Lon of observations, which I think is one of the worst errors. One history tape of station locations that we got has even more errors than are on the operational tapes; I think that no one had made time consistency checks on the position of stations. I think that enough of these problems can be solved in time to avoid major delays.

Can NMC reanalysis start on April 1?

There is still an awful amount of work to be done at NCAR and NMC, but I think we still have a chance to start in early April if we don't do some things that are less critical. TOVS is one of the problem areas, because NCAR does not have the NMC set yet, and it will take us some time to do merges. And this work detracts from other projects.

To answer the April 1 question, NCAR also needs to talk with NMC to scale the timing of the data tasks at NMC.

More details are in the following sections. This includes older data that can arrive at NMC at a later time.
4. **Rawinsondes**

Several main sets of global coverage:

a. NMC 1973-93
   - Is 1979-93 ready to send? Almost
   - What about station location errors?
   - See the more detailed notes at the end
b. NMC 1962-72. Will needs another 2 months on this.
   - Condense the aircraft format in this set
c. MIT set, 1958-63
d. TD54, 1948-on
   - GFDL has done most conversions, not all.
   - Talk to Will about conversions here
e. Review raob elevation problems
f. Do the text on old raobs
g. Our old AF synop raobs
   - Ignore for now
h. Five years of Coldfac data (maybe)
   - Talk with Bob Williams
i. There is some raob filler data from Japan

**Selected countries**

j. Do Australian update (should arrive soon)
k. Try to obtain Argentina (Kalnay has correspondence with them)
l. Get Brazil (asked on Dec 16 message to Brazil Jan 5)
m. S. Africa raobs (have 1968 through 1989)
n. Canada
   - Make corrections
o. Swiss
   - Write to them
p. China. NCAR has several years of early China data via E. Kung.
q. U.S. control raobs, 1946-on

**Other**

r. Line Islands
s. Gate raobs, 1974 (ships, drops, some land stations)
   - Bob will have these done in 2 weeks
t. S. African ships
u. German research ships
   - ECMWF wrote to them
   - Japan's Antarctic station. I've not been able to get it yet.
v. Russian Antarctic stations
w. Is any later U.S. Antarctic data ready?
x. Arctic raobs - Kahl
y. Russian GTS

And

z. Pick a set of remote raobs
aa. TOGA COARE raobs and drops
bb. Check on TD56 again (OL-A)
cc. Connections to Europe task

And other data

• Easter Island elevation
• Check with Shumbera on T3 raobs (old polar ice islands)
  — Jon Kahl is willing to do this key entry

And other issues

• Raob elevation and errors in raob locations

Upper winds - pibals

Do later. Some sets are at OL-A

5. Aircraft data

• This is coming along well
• Get the 1962-72 NMC data; Will Spangler is working on it
• See the other lists of these datasets

6. Cloud wind data

• See the text on lists in the Dec 10 reanalysis text
  — The cloud data on NMC UA tapes starts as early as 1967
  — We have NESS archives for 1974-on, but it's not up to date
  — We have GMS archives from Japan for 6 Apr 1978 - Dec 1991 (2 months are missing)
  — See the Dec 10 text about bias problems

7. Ice cap surface data

• Send the data for Antarctica and Greenland (it's from the University of Wisconsin)
8. Satellite sounder data

TOVS satellite sounding data (Nov 1978-recent)

- Need to do merges for best TOVS 2.5° archive (Nov 1978-on)
- NCAR needs the tapes from NMC
- And work with ECMWF to fill gaps
- The NASA Marshall group (Spencer and Christie) knows details about satellite-to-satellite TOVS microwave calibration changes that most other groups don't know.
- Other texts have much more information about TOVS

VTPR satellite sounding data (Nov 1972 - Feb 1979)

- NCAR could quickly extract the soundings when needed, or could just copy the whole dataset with radiances.

9. Surface 3-hour synop data

a. NMC, 1976-on
c. Global TD13 (1948-70), use 1948-66
d. Probably add U.S. hourly (1947-67), select 3-hour subset

Surface location and elevation

e. On the TD13 location data, maybe do diagnostic check for errors
f. AF station history - 700 pairs have errors (big moves, usually wrong). Figure it out.
   — Send a list to Bob Williams
g. The most recent station history is probably best, but there are real changes through time.
h. Did ECMWF compare surface locations over time?
i. Calculation of surface elevation? Has NMC or ECMWF done it?
   — Could do it from the analyzed surface pressure
j. Would like equations to calculate station-level pressure, given SLP and temp. This would need to include the plateau corrections vs. month that stations use. It's a hard task; skip it for now. Nearly all stations report SLP. This would allow people to calculate a station pressure when it isn't reported.

More details on GTS data from NMC follow:

10. NMC upper air data

There are 4 files:

- ADP-UPA (raobs and winds)
- aircraft
- cloud winds
- bogus (this file stopped Mar 11, 1986). The amount of data decreased starting about late 1984). The tape of bogus data from Australia should provide the necessary information.

Raob data are now in files for 0, 6, 12, and 18Z.

In early years they were grouped for 2 times per day, 0 and 12Z. The date 3 Feb 1975 was the earliest 6-hour time, then off and on. By 1 Mar 1975 it was always 6-hour times.

- We will discuss this change with NMC

When will NCAR send the NMC UA raob data to NMC (1973-92)

- The raobs from 1985 could be sent now
- The raobs from 1979 will be ready 15 Feb 1994
- Location of raobs. There are a few lat-lon locations that are extremely bad on the NMC tapes. NCAR will supply a correction file a little later for these stations.

Aircraft data

There is a 6-day period, Jul 20-25, 1985, when there were no 6 and 18Z files, but the data for these files was in the 0 and 12Z files.

- NCAR may fix these.

Sorts of data within a time period. From 1973-present, there have been three ways that the aircraft data has been sorted. NCAR will assume that the sort order within a file (time period) does not matter.

11. Surface 3-hour data from NMC (1976-on)

For earlier years through 1977, NMC put the 3-hourly data, as well as ship data, all on one tape series. There were 4 or 5 days on each tape received by NCAR. Then NMC used 2 tapes each week and put 0, 6, 12, and 18Z on one tape and 3, 9, 15, and 21Z plus data from all ships on the other tape.

NCAR has inventories that include the daily count of land stations each 3 hours. These inventories are on-line for 27 Nov 1988-on.

When NCAR sends this surface data to NMC, we will drop the ship data, because all of that plus more ships are in COADS.

- Shall NCAR put standard headers on these reports? Probably yes. It adds about 15% to the volume.
12. Rate of producing analyses at NMC

NMC plans to complete 35 years of analyses in about 2.5 years. This means that about 425 days of analyses must be completed each month (or 7.0 years of data done each 6 months). If NMC can achieve 30 days of analyses on each of 15 active days per month, then 450 days would be completed each month.

NMC will first analyze the data for 1985-93. NMC may be ready to start using 1979-84 data about Nov 1994.
Status of Reanalysis

Roy Jenne
NCAR
August 1994

- The text is mostly about the NMC/NCAR project to reanalyze global data for 40 years
- A little information is given about the status of the ECMWF project
- Data are output four times per day
- Data will be in both sigma and pressure coordinates
- Variables are listed: They include precipitation, snow, clouds, radiation, etc.
- There is a lot of data: Over 2000 Gbytes from the NMC/NCAR project
  - How will users get data access?
  - Can the costs be controlled?
Status of Reanalysis

This memo outlines the status of the NMC/NCAR reanalysis project. New global analyses will be made, 4 times a day, for a 40-year period, about 1956-95. The resolution is T62 (209 km), and there are 28 levels in the model. A little information about the ECMWF reanalysis is also included.

1. Order of doing reanalysis

NMC will first make new analyses for 1985-94, and then they will start on older data. They will be able to prepare a year of reanalyses during each 3 weeks.

2. Status on Friday, July 8, 1994 (from a talk with Kanamitsu at NMC)

The reanalysis of data for all of 1985 (which started with data for Dec 1984 for spinup) was completed recently. Statistics have been calculated and the data looks good. More checks will be made. It took NMC about 3 weeks to reanalyze the data for 1985. We will see below that some problems were found.

Kanamitsu says that they will probably start reanalyzing the data for 1986 today (July 8).

3. Problem with TOVS sounder data for early Jan 1986

On Thursday, July 14, 1994, Bob Kistler said that NMC has been working on the reanalysis of data for 1986 for about 6 days. When we talked, they had reanalyzed the data up to 4 March 1986. A few days earlier, when they ran the diagnostics on the analyses for Jan 1986, they found that the data was bad. They found that the reason for the problem was that TOVS data for the first 1 or 2 days of Jan 1986 was all messed up, so they had to reanalyze the data for Jan 1986 again.

4. Conference session on reanalysis

A forecasting conference was held in Portland, Oregon, July 18-22. There was a session on reanalysis. It also was a good time for people from NMC, NCAR, and ECMWF to talk with each other about the reanalysis projects.

About July 19, 1994, the NMC reanalysis was up to July 1986 (about 18 months done).

5. The Great Plains (U.S.) became too warm in summer 1985 (from a talk with Bob Kistler, NMC (1 Aug 1994)

July and August were too warm at the surface in the Great Plains, in the reanalysis for 1985. People at NMC looked at the whole annual cycle for temp and precip for the first year of reanalysis.
July and August were far too warm over the Plains—like 5°C too warm.

NMC found that a key reason is that the vegetation cover changes suddenly (from June to July) in the vegetation climatology they got from NASA. Bob Kistler noted that the annual temperature curve for a point in the U.S. Midwest looked almost like a step function in temp at the beginning of July. This reflects the harvesting of winter wheat; whole grid points change from a growing crop to dry stubble. The place then suddenly behaves almost like a desert.

The same vegetation climate is used in the NMC operational forecast runs. The floods in summer 1993 weren't handled very well. Kistler thinks that the vegetation climate is part of the reason.

Over the weekend (July 30, 31), Bob re-ran the summer 1985 data with different ways to handle vegetation. See the message from Eugenia Kalnay (in Attachment 4). She describes the problem and the plans in more detail.

There have been several cases where information learned from the reanalysis project has also improved the operational model at NMC. This is one such case. The new way to handle surface corrections was put into the production models about 11 Aug 1994.

6. When will NCAR obtain NMC/NCAR reanalysis data?

NMC will re-run the 18 months of reanalysis data they have done, using better surface conditions. NMC started the reanalysis again (about 15 Aug 1994), and they will be able to accomplish about 1.5 years of analysis each month. My guess is that NCAR will have the first 2 years of data by mid-Nov 1994, and the first 5 years of data within 3 more months.

7. Status of reanalysis in Aug 1994

On Aug 23, the analysis of data is now up to the end of May 1985, after the surface problem was fixed (NMC has done Dec 1984 to May 1985). On Aug 26, the analysis was near the end of Aug 1985.

- On Monday, Aug 29, NMC is doing Nov 1985 (so in 14 days they have done about 11 months).
- NMC will probably finish the 1985 through late 1993 period (9 years) about Jan 1995.

8. Sending observed data from NCAR to NMC (status on 15 July 1994)

- NCAR sent NMC all the types of data for 1985-89.
- Dennis Joseph sent data for 1990 to NMC on July 14.
- Data for 1991 was sent to NMC on July 18.

Some lost raobs have to be replaced: Sometime in 1991, NMC put a limit on the size of the buffer to save raob data, and the station catalog list was also involved. As time went on, a
lot of raobs were lost from many high block-station raobs (which also probably came in late). The loss gradually started about Jan 1992, and was fixed at NMC on about Jun 29, 1993. It especially affected the receipt of block 97 and 98 data. For example, about 30 good raob stations had almost no data during the first half of 1993 (and earlier) on these tapes. The problem was discovered at NCAR when people started looking for data to help the TOGA COARE experiment. NCAR is filling in the lost data from other sources.

9. What data will be available from NMC/NCAR reanalysis?

We will now summarize the data that will be available from the reanalysis project.

9.1 Primary data in sigma coordinates

The most basic output data is the atmospheric data in sigma coordinates, that includes levels in the earth’s boundary layer, and these levels follow a smoothed terrain. Also the flux fields on the Gaussian grid are part of the basic output.

- Summary of sigma coordinate archives and flux fields
  - Spectral T62, 28 sigma levels, each 6 hours. The volume is 2750 MB per year.
  - Sigma-level data on a Gaussian grid. It has precip, radiation, clouds, etc. The volume is 2377 MB per year.

Note: The lower sigma levels describe the world’s boundary layer, and these levels will be important inputs for various problems. In the 28-level model, the lowest few sigma levels are at 997.50, 992.16, 985.44, 976.27, 963.84, 947.13, 924.91, 895.78, and 858.33.

- More information about the primary data

There are two datasets in Attachment 1 that have all of the atmosphere dynamics in sigma coordinates data. These are:

- Item No. 6 (forecast). The volume is 2750 MB per year. This has initial conditions for 5-day forecasts. It is archived each 6 hours. It has spectral coefficients and flux fields, all as 32-bit numbers.
- Item No. 3 (GRIB analysis). The volume is 5064 MB per year. This file is on the Gaussian grid, in GRIB format. There are divergence and vorticity variables as well as wind components.
- Note that Item 6 has data, such as flux fields, that are in other files that are more compact. These should probably be removed from Item 6.

9.2 Data in pressure coordinates

- Synoptic order, 17 levels, 6 hourly. The volume is 2160 per year.
  - 17 levels, 1000 to 10 mb
  - Data each 6 hours, 2.5° grid
• Time series of data will also be available. About 1915 MB per year.
• Monthly means and other monthly statistics

9.3 The first guess fields

The first guess fields (from forecast models) that are used in the analyses will also be archived. We do not expect very many people to use these data:
— Guess in sigma coordinate (each 6 hours, 28 levels). The volume is 5068 MB per year.
— Guess fields transformed to pressure coordinate. The volume is 2160 MB per year.

9.4 Data on 11 potential temperature surfaces (6 hour, 2.5° grid)

There are 10 variables at each level. The volume is 2014 MB per year.

9.5 Data on CD-ROMs (one CD-ROM can hold 660 MB)

• A CD-ROM will be produced for every year of analyses
• Each will have a subset of pressure and isentropic level data
• Has many average fields
• Later CD-ROMs will have long time series for selected fields of data

9.6 Internal model diagnostics

These archives have such fields as internal model heating rates that are needed by only a few people.
• Full set. This has internal heating rates by radiation, condensation, etc. It has moistening and momentum terms. The data are at all levels every 6 hours. It is a big dataset. The volume is 9.70 GB per year.
• Reduced set. Monthly summary from the full set. The volume is only 188 MB per year.
• Question: Does a new type of reduced dataset need to be invented to reduce the volume of the full set to an intermediate size?

9.7 Plans for forecasts each few days

The present plan is to prepare a 5-day or an 8-day forecast each 5 days. TDL argues for an 8-day forecast, to get out to the point where there is no skill. The tentative plan is to prepare an archive of 2.5° pressure fields (17 levels), and the Gaussian flux fields every 12 hours. These would be like items 9 and 12 in Attachment 1.

What would the data volume be? With 8-day forecasts each 5 days, there would be 16 cycles of data each 5 days. This compares with four cycles each day in the archive of analyses. Therefore the archive would be 16/20, or 80% of the size of items 9 and 12 in Attachment 1. Therefore the volume of forecast data would be:

• Volume of pressure analysis: 80% of 2160 MB per year.
• Volume of flux fields: 80% of 1760 MB per year

10. **High-density tape drives at NMC**

NMC will obtain high-density tape drives on 30 Aug 1994 (goes from about 200 to 400 mb per cartridge). They have had 3480 drives until now (200 MB each). NCAR has had high-density tape drives for internal mass store use since Sep 1991. By Sep 20, 1994, we hope to have similar drives to use for import-export to read tapes from NMC.

11. **The ECMWF reanalysis** (status in Jul-Aug 1994)

*Status July 19:* ECMWF has completed about 6 months of reanalyses. They are doing the 15-year period, 1979-94. My guess is that the present work will be done again using 3D-var methods when that code is ready.

*Status Aug 25:* "Reanalysis is getting into gear nicely now. We had some initial problems with soil moisture, but I think we've overcome these now. Hope to have Jan through Nov of the FGGE year by the end of next month (Sep 1994).

The System is running well—we've gotten through 25 days in the last two working days!"

— from Rex Gibson, ECMWF

12. **The TOVS sounder data at NCAR**

The 2.5° TOVS data at NCAR includes the set saved by the research sections at NESDIS, with some gaps filled from the NCDC-Asheville archive. There are some gaps in our dataset of 2.5° TOVS sounder data. NASA has been able to fill most of the remaining gaps, based on NMC 1-per-day dump tapes that they have. Bob Kistler (NMC) will send us a copy of the data with 3 months of TOVS 2.5° gap fillers (in the original format). He is still waiting for another gap filler from Goddard.

13. **A new text about satellite sounder data** (for 25 years)

Jenne has nearly finished a new text about 25 years of satellite sounder data. It has information about TOVS, VTPR, and SIRS data. Data inventories are included. Some information about limb sounders and early stratospheric channels is included.

14. **SST and surface air temperature from NMC**

In the sigma coordinate system, the level of the ocean surface is not all at zero elevation. It has fixed waves that have significant amplitudes in some locations, such as near the Andes. NMC adjusts the observed SST so that the temperature difference between water temperature (SST) and air temperature is the same as it would be if the ocean surface were flat in the model (as it is in reality). The fields, such as 2m temperature over the ocean, will
include a small effect due to model elevation in some regions. The model elevation data is archived. Sea level pressure is valid at elevation equals zero.

15. About the data

The volume of various component datasets from NMC/NCAR reanalysis is in Attachment 1. The characteristics of different media that can be used to help distribute data is shown in Attachment 2. Briefing slides about different types of data from reanalysis are in Attachment 3. A memo from Kalnay is Attachment 4 (model surface conditions and NMC plans). A list of variables in different datasets is in Attachment 5.

16. The distribution of high-volume datasets

There are still big problems in distributing high-volume datasets, both because of tape handling and costs.

Some typical user costs from archives are:

- **NCAR**
  - $90 per tape, normal price (~$720 per Gbyte).
  - $70 per tape with volume discount (~$560 per Gbyte)

- **NCDC**
  - $180 per tape (~$1440 per Gbyte)
  - But some data is at a reduced price or given away to research users

The total output from reanalysis will be about 2000 Gbytes. This is equivalent to 16,000 high-density round tapes (6250 bpi)! Fortunately, most people will only use 5 to 30% of the data, but it still is a lot of data. It is clear that with normal prices most people could not afford the data; most people do not have a million dollars to spend.

NCAR has several large datasets where volume and cost are making distribution difficult. For about 2 years we have been developing strategies to cope with the problem. We will probably use Exabyte media, prepare a master tape copy that includes an efficient tape layout and error detection. Then we will automate the process of making copies (commercial firms may do this). A short text is almost available that describes these plans.

- A regular Exabyte tape holds 5 Gbytes and additional compression is possible.
- A longer Exabyte tape holds 7 Gbytes (started July 1994).
- A new technology Exabyte tape will hold 20 Gbytes (start Oct 1995).

For the many university users who compute at NCAR, they will also be able to access the data online.
17. Blocking the data onto tapes

A method of blocking the data will be used that can also be used for other types of data. The properties of the blocking method will be:

- There will be error protection on the data blocks.
- The reanalysis data will be kept in the GRIB format, it is now in this format. GRIB is an efficient format and an international standard.
- There will not be any slow software in the data path.
- The same software can be used to unblock other types of data.
- It will be possible to read the data on mainframes, workstations, and fast PCs.

18. Organization of the data

We have described several major types of data that are being output from reanalysis. The data will be put on tapes according to these categories, so that people can obtain one type of data without bothering with all of the rest. If people order whole types of data, the price will be much less than if data selection is required.

- End -
Last Friday I created the volsets for storing the Reanl output for 1985. First I extrapolated from a set of files for one month the total blocks for each volset I planned to create for a year. Then I divided by 6700, an estimate of blocks/3480 and rounded up. This exercise produced a count of 262 volumes. I did not have scratch volumes available, so I returned to pool ri23bk/reanl volumes from volsets no longer in use. While this did not produce what I ultimately would want (a contiguous sequence), it would permit me to test out the scripts to predict the volume count for a year.

The following table (file /reanl1/ri2/volset_tapes_1985) summarizes that process (the actual volumesets have 1985 appended to the name).

<table>
<thead>
<tr>
<th>Volset</th>
<th>Blocks</th>
<th>Format</th>
<th>Output</th>
<th>3480's</th>
<th>MB/yr</th>
<th>GB/year</th>
<th>Type of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>grb3d</td>
<td>grib</td>
<td>primary</td>
<td>45</td>
<td>9701</td>
<td>388</td>
<td>3D heating, momentum, moisture</td>
</tr>
<tr>
<td>2</td>
<td>prepqm</td>
<td>bufr</td>
<td>primary</td>
<td>34</td>
<td>7187</td>
<td>287</td>
<td>observe data with all metadata</td>
</tr>
<tr>
<td>3</td>
<td>grbsanl</td>
<td>grib</td>
<td>primary</td>
<td>24</td>
<td>5064</td>
<td>203</td>
<td>sigma analysis</td>
</tr>
<tr>
<td>4</td>
<td>grbsfo6</td>
<td>grib</td>
<td>primary</td>
<td>24</td>
<td>5068</td>
<td>203</td>
<td>sigma quasi</td>
</tr>
<tr>
<td>5</td>
<td>recv</td>
<td>binary</td>
<td></td>
<td>22</td>
<td>1276</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>fcst</td>
<td>binary</td>
<td></td>
<td>21</td>
<td>2750</td>
<td>110</td>
<td>initial codes for 5-day forecasts</td>
</tr>
<tr>
<td>7</td>
<td>2d</td>
<td>grib</td>
<td>reduced</td>
<td>12</td>
<td>2377</td>
<td>95</td>
<td>has &gt;7 files/tape plus more</td>
</tr>
<tr>
<td>8</td>
<td>theta</td>
<td>grib</td>
<td>reduced</td>
<td>10</td>
<td>2014</td>
<td>81</td>
<td>110 surfaces, 11 vel/each</td>
</tr>
<tr>
<td>7</td>
<td>pgb.f00</td>
<td>grib</td>
<td>primary</td>
<td>10</td>
<td>2160</td>
<td>86</td>
<td>pressure analyses</td>
</tr>
<tr>
<td>6</td>
<td>pgb.f06</td>
<td>grib</td>
<td>primary</td>
<td>10</td>
<td>2160</td>
<td>86</td>
<td>pressure quasi</td>
</tr>
<tr>
<td>11</td>
<td>prs</td>
<td>grib</td>
<td>reduced</td>
<td>9</td>
<td>1915</td>
<td>77</td>
<td>pressure time series</td>
</tr>
<tr>
<td>12</td>
<td>grb2d</td>
<td>grib</td>
<td>primary</td>
<td>9</td>
<td>1760</td>
<td>70</td>
<td>the fluxes</td>
</tr>
<tr>
<td>13</td>
<td>ccc</td>
<td>ascii</td>
<td>primary</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>internal use only</td>
</tr>
<tr>
<td>14</td>
<td>hpvnl</td>
<td>grib</td>
<td>primary</td>
<td>7</td>
<td>1416</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>ieex</td>
<td>grib</td>
<td>reduced</td>
<td>2</td>
<td>457</td>
<td>18</td>
<td>no worry, use for G/C</td>
</tr>
<tr>
<td>16</td>
<td>oiqc</td>
<td>ascii</td>
<td>primary</td>
<td>3</td>
<td>440</td>
<td>18</td>
<td>dupes of metadata</td>
</tr>
<tr>
<td>17</td>
<td>print</td>
<td>ascii</td>
<td>primary</td>
<td>3</td>
<td>576</td>
<td>23</td>
<td>signotes</td>
</tr>
<tr>
<td>18</td>
<td>cdm</td>
<td>grib</td>
<td>reduced</td>
<td>1</td>
<td>188</td>
<td>108</td>
<td>from 3A People use this</td>
</tr>
<tr>
<td>19</td>
<td>3d</td>
<td>grib</td>
<td>reduced</td>
<td>1</td>
<td>80</td>
<td>3</td>
<td>diagnostics, may be of interest</td>
</tr>
<tr>
<td>20</td>
<td>znl</td>
<td>binary</td>
<td>primary</td>
<td>1</td>
<td>80</td>
<td>7</td>
<td>no means fields, etc, on 2.5</td>
</tr>
<tr>
<td>21</td>
<td>grib</td>
<td>grib</td>
<td>reduced</td>
<td>1</td>
<td>182</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>misc</td>
<td>binary</td>
<td>reduced</td>
<td>1</td>
<td>129</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>daily</td>
<td>grib</td>
<td>reduced</td>
<td>1</td>
<td>214</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>sigma</td>
<td>binary</td>
<td>reduced</td>
<td>1</td>
<td>22</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>optavg</td>
<td>binary</td>
<td>primary</td>
<td>1</td>
<td>33</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Total volume (not "mo") = 1854 GBytes

* Theta has grown (more levels, etc.) and it is in 2 forms. The big set is "theta" above. What is 2nd one?

* Main analyses for use: 3, 7, 8, 9, 11; volume is 542 GBytes for 40 yrs

* 3D heating etc (19, 19?) What in description? Can must use #19?

* binary means 32-bit numbers
## Reanalysis Archives

<table>
<thead>
<tr>
<th>Type</th>
<th>Grid</th>
<th>Per Day</th>
<th>Bits per No.</th>
<th>No. of Levels</th>
<th>Fields at Time</th>
<th>Gbytes 40 Years</th>
<th>Popularity (0 to 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigma anal</td>
<td>Spect</td>
<td>4</td>
<td>28</td>
<td>6*28</td>
<td>110</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Sigma anal (stack)</td>
<td>Full</td>
<td>4</td>
<td>10</td>
<td>28</td>
<td>6*28</td>
<td>203</td>
<td>6 (low levels)</td>
</tr>
<tr>
<td>Theta (11 levels)</td>
<td>2.5</td>
<td>4</td>
<td>10</td>
<td>11</td>
<td>11*10</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>Pressure anal (stack)</td>
<td>2.5</td>
<td>4</td>
<td>9</td>
<td>17</td>
<td>85</td>
<td>86</td>
<td>9</td>
</tr>
<tr>
<td>Pressure, time series</td>
<td>2.5</td>
<td>4</td>
<td>9</td>
<td>17</td>
<td>60?</td>
<td>77</td>
<td>10</td>
</tr>
<tr>
<td>Sigma guess (stack)</td>
<td>Full</td>
<td>4</td>
<td>10</td>
<td>28</td>
<td>6*28</td>
<td>203</td>
<td>0.2</td>
</tr>
<tr>
<td>Pressure guess (stack)</td>
<td>2.5</td>
<td>4</td>
<td>9</td>
<td>17</td>
<td>85</td>
<td>86</td>
<td>0.2</td>
</tr>
<tr>
<td>Fluxes (precip, etc.)</td>
<td>Full</td>
<td>4</td>
<td>8</td>
<td>34</td>
<td>70</td>
<td>6 to 10</td>
<td></td>
</tr>
<tr>
<td>Fluxes (CAC subset)</td>
<td>Full</td>
<td>4</td>
<td>8</td>
<td>24</td>
<td>22</td>
<td>6 to 10</td>
<td></td>
</tr>
<tr>
<td>Diagnostic stack, all</td>
<td>Full</td>
<td>4</td>
<td>9</td>
<td>17</td>
<td>5*17</td>
<td>388</td>
<td>0.05</td>
</tr>
<tr>
<td>Diagnostic (mo stat)</td>
<td>2.5</td>
<td>9</td>
<td>17</td>
<td>150?</td>
<td>8</td>
<td>8</td>
<td>0.8</td>
</tr>
<tr>
<td>Set of monthly stat</td>
<td>2.5</td>
<td></td>
<td>7</td>
<td>7</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: A global grid has 192*94 points (Gaussian)
— The resolution (T62) is 1.88° or 209 km
A global 2.5° grid has about 145*73 points

Roy Jenne
July 1994
Attachment 2

Distribution of Data on Tapes

The lowest cost method to distribute large amounts of data is usually to use tapes. The cost to copy tapes is significant, but it can be controlled. Table 1 has some information about the evolution of selected tape capability and cost. The more recent IBM and Exabyte tape drives have a data compression option built into the drives. The numbers in Table 1 are without compression.

Table 1. Magnetic Tape Information for 1960-95

<table>
<thead>
<tr>
<th>Media Holds (in MB)</th>
<th>Media Cost per GB</th>
<th>Tapes per 100 GB</th>
<th>Tape Compress</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Round half-inch tapes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960 (7-track, 556 bpi, $13)</td>
<td>12</td>
<td>$1100</td>
<td>8330</td>
<td>No</td>
</tr>
<tr>
<td>1972 (9-track, 1600 bpi, $10)</td>
<td>37</td>
<td>270</td>
<td>2700</td>
<td>No</td>
</tr>
<tr>
<td>1980-94 (9-track, 6250 bpi, $10)</td>
<td>125</td>
<td>80</td>
<td>800</td>
<td>No</td>
</tr>
<tr>
<td>2. IBM 3480, 3490 cartridges</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb 1986 (200 MB, $14)</td>
<td>174</td>
<td>81</td>
<td>575</td>
<td>No</td>
</tr>
<tr>
<td>Sep 1989 (230 MB, $5)</td>
<td>200</td>
<td>25</td>
<td>500</td>
<td>No</td>
</tr>
<tr>
<td>May 1991 (460 MB, $5)</td>
<td>410</td>
<td>12.2</td>
<td>244</td>
<td>Yes</td>
</tr>
<tr>
<td>3. Exabyte</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988 8 mm (2.2 GB, $10)</td>
<td>2100</td>
<td>4.76</td>
<td>48</td>
<td>No</td>
</tr>
<tr>
<td>1991 8 mm (5.0 GB, $9)</td>
<td>4700</td>
<td>1.92</td>
<td>21.3</td>
<td>Yes</td>
</tr>
<tr>
<td>Jul 1994 8 mm (7.0 GB, $10)</td>
<td>6700</td>
<td>1.49</td>
<td>14.9</td>
<td>Yes</td>
</tr>
<tr>
<td>Oct 1995 8 mm (20 GB $20?)</td>
<td>1900</td>
<td>1.05</td>
<td>5.3</td>
<td>Yes</td>
</tr>
<tr>
<td>4. IBM NTP tape drives</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall 1994, NTP (10 GB)</td>
<td>9700</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>5. CD-ROM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986-94</td>
<td>660</td>
<td>5</td>
<td>152</td>
<td>No</td>
</tr>
</tbody>
</table>

R: for a 2x speed CD-ROM drive

Cost of drives: The prices of drives (the best prices) are about $12,000 for cartridges; $2100 for Exabyte tapes; $200 for CD-ROMs.

* These are data rates for a relatively fast drive that moves tapes at 150 inches per second.
Reanalysis Archives

by:
John E. Janowiak, Wesley Ebisuzaki, Muthuvel Chelliah, Suranjana Saha, Robert Kistler, Masao Kanamitsu and Glenn White

National Meteorological Center
NOAA/National Weather Service

* Presented by Wesley at Forecast Conference July 1994

Something for Everyone:

(1) Global Change/Long-term Trends
   - gridded analysis by 1 system, past and future analyses
   - not free of all artificial trends, changes in observing network
     Two approaches:
     (1) try to get best analysis
     (2) keep observing network uniform in time
   - needs: long homogenous time series with few variables

(2) Low Frequency Variability, General Circulation Studies
   - uniform/state-of-art analysis system
   - longer record (NMC global analysis starts in 1978)
   - more vertical resolution (17 prs levels)
   - more variables (heating, omega, etc.)
   - pressure, theta, sigma vertical coordinates
   - some terms for angular momentum budgets
   - stratosphere (28 model levels)
   - 6 hour resolution
(3) Radiation Studies
- basic fluxes at the top of the atmosphere, surface
- heating rates (by SW, LW) on sigma levels
- Cloud forcing (type II)
- can be compared with satellite data
- Clear sky fluxes

(4) Numerical Modellers
- some variables that only a numerical modeller would love
- more of a diagnostic for the forecast model rather than an analyzed quantity, useful for further development
- 99% of the variables calculated in the AMIP experiment (Atmos. Model Intercomparison Project)
- AMIP-like simulation using same forecast model

(5) Others
- wind stress over the ocean (Coupled Models)
- energy/water budget at surface

---

**ISENTROPIC COORDINATES**

- new

- 11 levels
  - 270 K
  - 280 K
  - 290 K
  - 300 K
  - 315 K
  - 330 K
  - 350 K
  - 400 K
  - 450 K
  - 550 K
  - 650 K

- variables
  - Absolute vorticity
  - Potential vorticity
  - Montgomery stream-function
  - Brunt-Vaisala Frequency/2
  - Relative Humidity
  - Temperature
  - \( \Theta \)
  - \( U, V \)
  - mass-weighted \( U, V \) (i.e. \( U^*d\Theta/d\Theta \))
  - \( \Theta \) at surface

- 6 hour, 25 degree x 25 degree grid
- monthly means/variances

---

**Pressure Coordinates**
- 17 levels (new: 600 mb, 925 mb, stratosphere)
  - 10, 20, 30, 50, 70, 100, 150,
  - 200, 250, 300, 400, 500, 600,
  - 700, 850, 925, and 1000 mb

**6 Hours Time series**
- \( U, V \)
- Geopotential height
- Temperature
- Relative/Specific Humidity (to 300 mb)
- \( \omega \) (to 100 mb)

**Monthly Means**
- above
- divergence, velocity potential
- vorticity, stream function
- \( UT, VT, \omega T \)
- \( Uq, Vq, \omega q \)
- \( UU', UV, U\omega', V\omega', VV, \omega\omega \)
- virtual temperature

---

**Zonal Cross Sections**
( Monthly mean )

**Pressure surfaces**
- Geopotential Height *
- Temperature *
- \( U, V \)
- \( dP/dt \)
- Relative, specific humidity *
- Meridional stream function
- \( UV \)
- \( Vq' \)
- \( VT \)

**Theta surfaces**
- \( N^2 \)
- potential vorticity
- pressure
- relative humidity
- \( U, V \)
- mass-weighted \( U, V \)

* variance and monthly mean
sigma

- For high-accuracy calculation, need to use data on sigma surfaces (ex. difficult to resolve boundary layer with standard pressure surfaces)

- Two versions of sigma file:
  - GRIB 2.5 degree x 2.5 degree (3.4 MB/6 hours)
  - Spectral T62, (1.8 MB/6 hours)

- Three products calculated from Sigma files:
  - Moisture Budget (J. Wang)
  - Monthly mean Sigma file
  - Various Angular momentum terms (with AER)

- Variables include:
  - U, V
  - divergence, vorticity
  - Tw, q
  - ln(surface pressure)

  Diabatic heating terms
    - large-scale condensation
    - deep/shallow convection
    - vertical diffusion
    - short wave
    - long wave

  Vertical diffusion of U/V momentum

Surface Quantities

Radiative
  - surface albedo (given)
  - skin temperature
  - cloud forcing (type II)
    - long/short wave, top/surface
  - fluxes at surface/top of atmosphere
  - clear sky fluxes at sfc/top of atmosphere

Surface Conditions
  - ice (given)
  - skin temperature/SST
  - snow depth
  - runoff
  - potential evaporation
  - soil moisture (2 levels)

Atmospheric Variables

  - surface pressure
  - mean sea-level pressure
  - total, convective precipitation
  - precipitable water
  - surface stresses
  - 10 m U/V
  - 2 m T, q, RH
  - cloud amounts (high/low/medium)
  - gravity wave drag (u/v stress)
  - sensible, latent heat fluxes
- need resolution between 6 hours and 1 month
- some monthly mean files are striated by hour
  ex. 00Z precipitation
- some fields (24 hour average)
  total, convective precipitation
  \( U, V, T \), height: 200, 500, 700, 850, 925 mb
  net short/long wave flux at surface
  sensible, latent heat flux
  OLR
  surface stress
  2 m temperature
  specific humidity 925, 850, 700, 500 mb

- 00Z time series
  potential vorticity, \( U, V, T \) at 315K, 350K
  skin temperature
  soil moisture at 5, 50 cm
  geopotential height, \( T \) at 100, 70 and 50 mb

- 00Z, 12Z time series (twice daily)
  height, \( T, U, V \) at 200, 500, 850 mb
  precipitable water
  specific humidity, temperature at 2 m
  \( \omega \) at 500 mb
  \( U, V \) at 10 m
  surface pressure
  1000 mb height.

Distribution

CD-ROM:
1 CD-ROM for every year of analyses
subset of pressure data, isentropic data
(limited vertical resolution, limited num. variables)
many average fields (mean, variances, covariances)
1 CD-ROM for time series (after N years)
very limited number of variables/levels

Anonymous FTP: nic.fb4.noaa.gov
15 GB reserved for reanalysis products
keep a limited number of time series on-line (?)
keep last year of reanalysis on-line
people must ftp the data before the next year's analysis
replaces the older analysis
might get a faster transfer if you use Nic at Nite

TAPE:
Sent to NCAR, NCDC (Nat. Climate Data Center), CDC
Requests for tapes have to directed to NCAR, NCDC or CDC
NMC's official policy is not to supply tapes to outside users
(not our job, no money).

Other:
Unofficial data sites

Summary

The reanalysis project at NMC has started.

The results of reanalysis can be used by a larger
community than the previous system
- homogenous system
- more analyzed fields
- 6 hr resolution
- longer period

Reanalysis (past) \( \rightarrow \) CDAS (current)
past and monitoring the present.

Tried to make the data easy to use (GRIB/IEEE) and
accessible.

With reanalysis, studies that could not have been
previously attempted are now possible.

handout: proposal for CD-ROM distribution
suggestions welcome
Date: Tue, 9 Aug 94 18:26:54 EDT
From: Eugenia Kalnay <wd23ek@sun1.wwb.noaa.gov>
To: deparker@email.meto.govt.uk
Subject: GISST and Reanalysis
Cc: ao@gfdl.gov, jnp@mines.utah.edu, jenne@ncar.ucar.edu,
    ken@msmail.ncdc.noaa.gov, coughlan@ogp.noaa.gov, mooney@ogp.noaa.gov,
    mcperson@sun1.wwb.noaa.gov, wallace@atmos.washington.edu,
    err@ecmwf.co.uk, schubert@schubert.gsfc.nasa.gov,
    mcane@ldgo.columbia.edu, shukla@cola.iges.org

Thank you for your letter of 20 July. Before answering, let me give you an update on where we are:

- Problem in simulation of plants -

We started the Reanalysis in June and the system we designed behaved as we were hoping, we were able to do two years (1985 and 1986) at a rate of a year in less than 3 weeks. We stopped to look at the results, and they seemed generally very reasonable except for too high T and low pp over the midwest US in July and August. It turned out that we were using a climatology of plant resistance created by SiB that was far too high over the US in the summer (almost 1000, compared to a typical value of about 50). So we have tested new fields for resistance and albedo, and are getting much better results. Naturally there are other problems also (e.g., surface solar radiation flux not as good as we would want), but those we cannot fix quickly, so they will have to wait for the second reanalysis.

We have decided to restart the reanalysis in a few days, and still plan to complete the 1985-1993 period this year. Next we'll do 1979-1984, followed by 1958-1978, and hopefully 1946-1957. We should finish all 40+ years in less than 3 years. We also decided, in response to several suggestions, including the UKMO, to save one year rather than two years per CDrom, so that we can put further vertical levels and fields. We should start producing them later this year.

In view of this schedule, we won't need the GISST data before we do 1979-1984 (for the period before 1982), and therefore we don't need it until early next year. Gisst 2.2 sounds from your description better than gisst 2.1, so we would be probably very grateful if you could provide it to us.

Please pass on this message to anyone interested in the project. We really appreciate you collaboration very much!

Sincerely, Eugenia Kalnay
Chief, Development Division
National Meteorological Center
Washington DC 20233
tel: 301-763-8005
fax: 301-763-8545
email: wd23ek@sun1.wwb.noaa.gov
or e.kalnay/omnet

* NMC will use SST and sea ice grids from the UK for the earlier years
Variables in different files (NHC/NCAR Reanalysis Project)

All files are 4 times a day at 00, 06, 12 and 18 GMT

1. Pressure file (analysis and 6 hour forecast guess on 2.5 degree latitude-longitude grid)

Geopotential height (gpm) at 16 levels
u-wind at 16 levels (m/s)
v-wind at 16 levels (m/s)
Relative humidity at lowest 7 levels (%)
Temperature (thermodynamic not virtual) at 16 levels (K)
Pressure vertical velocity at lowest 11 levels (Pa/s)
Surface pressure (Pa)
Surface skin temperature (K)
Pressure vertical velocity at the surface (Pa/s)
Relative humidity at the surface (%)
Temperature at tropopause level (K)
Pressure at tropopause level (Pa)
u wind at tropopause level (m/s)
v wind at tropopause level (m/s)
Vertical speed shear at tropopause level (ls)
Best (4-layer) lifted index (K)
Pressure at maximum wind level (Pa)
u wind at maximum wind level (m/s)
v wind at maximum wind level (m/s)
Surface geopotential height (gpm)
u wind at the lowest model level (m/s)
v wind at the lowest model level (m/s)
Potential temperature at the lowest model level (K)
Layer mean relative humidity (sig=1-0.40, 0.93-0.66, 0.66-0.40) (%)
Precipitable water (kg/m³)

Standard pressure levels: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30
20,10 hPa

All parameters are instantaneous values

Note: This list was from Masao Hanamura (NHC) on
26 Aug 1994

Roy Jenera, NCAR
2. Surface flux file (On the model Gaussian grid 192 x 64)

Zonal wind stress (N/m²) (A)
Meridional wind stress (N/m²) (A)
Sensible heat flux (W/m²) (A)
Latent heat flux (W/m²) (A)
Skin temperature (K) (I)
Soil moisture content (kg/m³) (I)
Water equivalent of accumulated snow depth (kg/m³) (I)
Downward short wave radiation flux at the top (W/m²) (A)
Upward short wave radiation flux at the top (W/m²) (A)
Upward long wave radiation flux at the top (W/m²) (A)
Downward long wave radiation flux at the surface (W/m²) (A)
Upward long wave radiation flux at the surface (W/m²) (A)
Downward short wave radiation flux at the surface (W/m²) (A)
Upward short wave radiation flux at the surface (W/m²) (A)
Total high level cloud cover (%) (A)
Nearby model level of high cloud top (int) (A)
Nearby model level of high cloud bottom (int) (A)
High cloud top temperature (K) (A)
Total middle level cloud cover (%) (A)
Nearby model level of middle level cloud top (int) (A)
Nearby model level of middle level cloud bottom (int) (A)
Middle cloud top temperature (K) (A)
Total low level cloud cover (%) (A)
Nearby model level of low cloud top (int) (A)
Nearby model level of low cloud bottom (int) (A)
Low cloud top temperature (K) (A)
6 hr total precipitation (kg/m³) (Aabh)
6 hr convective precipitation (kg/m³) (Aabh)
Ground heat flux (W/m²) (A)
Land-sea mask (1=land; 0=sea; 2=sea-ice) (I)
u wind at 10 m (m/s) (I)
v wind at 10 m (m/s) (I)
Temperature at 2 m (K) (I)
Specific humidity at the lowest model level (kg/kg) (I)
Surface pressure (Pa) (I)

(A) .. Average in 6 hr forecast
(I) .. Instantaneous value at forecast hour 6
(Aabh) .. Accumulation in 6 hour forecast
3. Diagnostic file at FT=6 hr (on the model Gaussian grid 192 x 94)

Surface skin temperature (K) (I)
Soil moisture content (kg/m²) (I)
Water equivalent of accumulated snow depth (kg/m²) (I)
Soil temperature at 10, 50 and 500 cm (K) (I)
Surface roughness (m) (I)
Total 6 hour accumulated precipitation (kg/m²) (A, B, C)
Convective 6 hr accumulated precipitation (kg/m²) (A, B, C)
Sensible heat flux (W/m²) (A)
Zonal wind stress (N/m²) (A)
Meridional wind stress (N/m²) (A)
Latent heat flux (W/m²) (A)
Albedo (%) (I) *
Downward short wave radiation flux at the top (W/m²) (A)
Upward short wave radiation flux at the top (W/m²) (A)
Upward long wave radiation flux at the top (W/m²) (A)
Downward long wave radiation flux at the surface (W/m²) (A)
Upward long wave radiation flux at the surface (W/m²) (A)
Downward short wave radiation flux at the surface (W/m²) (A)
Upward short wave radiation flux at the surface (W/m²) (A)
Total high level cloud cover (%) (A)
Nearby model level of high cloud top (int) (A)
Nearby model level of high cloud bottom (int) (A)
Total middle level cloud cover (%) (A)
Nearby model level of middle level cloud top (int) (A)
Nearby model level of middle level cloud bottom (int) (A)
Total low level cloud cover (%) (A)
Nearby model level of low cloud top (int) (A)
Nearby model level of low cloud bottom (int) (A)
Convective cloud cover (%) (A)
Clear sky upward long wave flux (W/m²) (A)
Clear sky upward solar flux (W/m²) (A)
Clear sky downward long wave flux (W/m²) (A)
Clear sky downward solar flux (W/m²) (A)
Clear sky upward solar flux (W/m²) (A)
Cloud forcing net solar flux (W/m²) (A)
Cloud forcing net solar flux at the surface (W/m²) (A)
Cloud forcing net solar flux (W/m²) (A)
Cloud forcing net long wave flux (W/m²) (A)
Cloud forcing net long wave flux (W/m²) (A)
Visible beam downward solar flux at the surface (W/m²) (A)
Visible diffuse downward solar flux at the surface (W/m²) (A)
Near IR diffuse downward solar flux at the surface (W/m²) (A)
Large scale condensation heating rate at 28 model levels (K/s) (A)
Deep convective heating rate at 28 model levels (K/s) (A)
Deep convective moistening rate at 28 model levels (Kg/Kg/s)
Shallow convective heating rate at 28 model levels (K/s)
Shallow convective moistening rate at 28 model levels (Kg/Kg/s)
Vertical diffusion heating rate at 28 model levels (K/s)
Vertical diffusion zonal acceleration at 28 model levels (m/s²)
Vertical diffusion meridional acceleration at 28 model levels (m/s²)
Vertical diffusion moistening rate at 28 model levels (Kg/Kg/s)
Solar radiative heating rate at 28 model levels (K/s)
Long wave radiative heating rate at 28 model levels (K/s)

Note that some parameters are redundant with surface flux file

4. Diagnostic file at FT=0 (on model Gaussian grid 192 x 94)
   Surface skin temperature (K) (I)
   Soil moisture content (kg/m²) (I)
   Water equivalent of accumulated snow depth (kg/m²) (I)
   Soil temperature at 10, 50 and 500 cm (K) (I)
   Surface roughness (m) (I) *

(I) Instantaneous values
*) Not a predicted parameter but changes because of the time interpolation of monthly climatology

5. Grib sigma file (on model Gaussian grid 192 x 94)
   Relative vorticity at 28 model levels (1/s)
   Divergence at 28 model levels (1/s)
   Thermodynamic temperature (not virtual) at 28 model levels (K)
   Specific humidity at 28 model levels (kg/kg)
   x-gradient of log surface pressure (1/m)
   y-gradient of log surface pressure (1/m)
   u wind (m/s) at 28 model levels
   v wind (m/s) at 28 model levels
   Surface pressure (Pa)
   Surface geopotential height (gpm)
   x-gradient of height (m/m)
   y-gradient of height (m/m)

6. Zonal diagnostic file
   Multi-level fields divided into the area 90N-60N, 60N-30N, 30N-30S, 30S-60S and 60S-90S.
Global averages are also computed.

- u component of winds (I)
- v component of winds (I)
- Virtual temperature (I)
- Specific humidity (I)
- Squared vorticity (I)
- Squared divergence (I)
- Pressure vertical velocity (I)
- Temperature (I)
- Relative humidity (I)
- Kinetic energy (I)
- Convective heating rate (A)
- Large scale heating rate (A)
- Shallow convection heating rate (A)
- Heating due to vertical diffusion (A)
- Convective moistening rate (A)
- Shallow convection moistening rate (A)
- Moistening due to vertical diffusion (A)
- u-acceleration by vertical diffusion (A)
- v-acceleration by vertical diffusion (A)
- Heating by short wave radiation (A)
- Heating by long wave radiation (A)

Single level field separated by surface characteristics i.e. snow covered land/sea-ice/snow-covered-sea-ice/sea over the same latitudinal belt as for the multi-level fields

- Total precipitation (Acc)
- Convective precipitation (Acc)
- Sensible heat flux (A)
- Latent heat flux (A)
- u-stress (A)
- v-stress (A)
- Area coverage of total rain (Acc)
- Area coverage of convective rain (Acc)
- Surface pressure (I)
- Skin temperature (I)
- Soil wetness (I)
- Snow depth (I)
- 10 cm deep soil temperature (I)
- 50 cm deep soil temperature (I)
- 500 cm deep soil temperature (I)*
- Surface net short wave radiation flux (A)
- Surface net long wave radiation flux (A)
- Relative humidity at the lowest model level (I)
- Virtual temp at the lowest model level (I)
Temperature at the lowest model level (I)
Specific humidity at the lowest model level (I)
Surface roughness (I) *
Land/sea/sea-ice mask (I) *
U-acceleration by gravity wave drag (A)
V-acceleration by gravity wave drag (A)
Angular momentum budget surface torque (A)
Angular momentum budget gravity wave drag (A)
Angular momentum mountain torque (A)
Total angular momentum (I)
Planetary angular momentum (I)

(A) .. Average in 6 hr forecast
(I) .. Instantaneous value at forecast hour 6
(Acc) .. Accumulation in 6 hour forecast
*) Not a predicted parameter but changes because of the time interpolation of the monthly climatology and snow cover change

7. Restart file (no grib packing is done to insure reproducibility).

7.1 Sigma file (T62 L28) in spectral coefficient form

Vorticity at 28 model levels (1/s)
Divergence at 28 model levels (1/s)
Virtual temperature at 28 model levels (K)
Specific humidity at 28 model levels (Kgm/Kgm)
Log of surface pressure
Surface geopotential

7.2 Surface file (Gaussian grid 192 x 94)

Surface temperature
Soil moisture
Snow depth water equivalent
10 cm deep soil temperature
50 cm deep soil temperature
500 cm deep soil temperature

Surface roughness
Convective cloud cover
Convective cloud bottom height
Convective cloud top height
Land/sea/sea-ice mask
Plant evaporation resistance
Ratio between 10 m winds and lowest model level winds
Considerations for Reanalysis

Changes in the input data coverage will affect what happens in reanalysis. This text reminds people of some of the issues that we should think about.

1. Reanalysis of TOGA COARE period

The main months of data for the TOGA COARE experiment were about Oct 1992 through Feb 1993.

I hope that reanalysis will at least include the additional raob sites, and the research aircraft data.

2. Ship and buoy data for 1978-79

In a few months the COADS project (by NCAR, ERL, NCDC) will have newer data for 1978-79, and for earlier years also.

We understand that reanalysis at both NMC and ECMWF will use FGGE data for the 12-month period (Dec 1978 - Nov 1979). COADS will be used for the other months. Remember that two fixed buoys had wildly wrong locations for a month or so during FGGE. Information about necessary corrections was given to ECMWF.

3. Satellite sounder data in 1978-79

The dates of sounder data follow:

SIRS: Apr 1969 - Sep 1972
VTPR: Nov 1972 - 28 Feb 1979
TOVS (2.5°): 1 Jan 1979 - 19 Jan 1994 (24 GB)
TOVS (raw): 29 Oct 1978-on

- I hope that reanalysis can obtain benefit from TOVS starting 29 Oct 1978 (we would need to prepare 2.5° TOVS for 2 months before the present 2.5° data starts on 1 Jan 1979).

- Overlap period between VTPR and TOVS (Nov 1978 - Feb 1979). The overlap period will permit some tests between the two sounders.

- The data counts for SIRS are sometimes low.

4. Will analyses methods get lost in S. Hemisphere, in early years?

From late 1972-on, we have satellite sounder data that will help to keep the analyses in control over the S. Hemisphere. Before that, there is some Nimbus SIRS sounder data for a few years. Without sounder data, or bogus help, there may not be enough data to keep
major troughs in the right place. In this case the forecast guess would also be bad. Good observations could easily be rejected because of a bad guess. If this starts to happen, is there any way to avoid it?

One option is to use some of the available daily analyses for the S. Hemisphere (at NCAR) to generate bogus data:

- Analyses, 1000-100 mb from Australia (Apr 1972-present)
- S. Hemisphere, 500 mb height (Aug 1968 - Jun 1977), from Australia
- Part of S. Hemisphere (60-210°E), SLP (Jun 1957 - Feb 1978), from New Zealand
- IGY, 18 months of S. Hemisphere, SLP and 500 mb height (Jun 1957 - Dec 1958)

<table>
<thead>
<tr>
<th></th>
<th>1960</th>
<th>1970</th>
<th>1980</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>early sfc analyses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td>S. Hemisphere 500 mb</td>
<td>Apr 1968</td>
<td></td>
</tr>
<tr>
<td>1957</td>
<td>40% of Hemisphere</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SLP from New Zealand</td>
<td>Apr 1972</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IGY SLP and 500 mb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td>S. Africa drew maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1957</td>
<td>Dec 1958</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Data near 150 mb for S. Hemisphere

There were two periods when there were many constant level balloons over the S. Hemisphere:

- DS800   EOLE balloons, Aug 1971 - Dec 1972, from France
- DS615   TWERLE, Jun 1975 - Aug 1976

The EOLE balloon data only gives time, location and winds. The TWERLE balloons also had a very accurate altimeter that measured the distance to the ocean surface. Suppose that a balloon is at 162 mb. We will then know the height of the 162-mb surface plus the wind. This is like having one mandatory level in a rawinsonde. I understand that such a single level raob cannot be fully used now, but in principle it could be used. Now it is used like an aircraft wind report.

6. Duplicate raobs; pick the best or combine them

Suppose that for reanalysis, we have duplicate rawinsondes at a site for a given time. One observation may come from GTS and another may come from a national archive. There are other sources for duplication for reanalysis. Most methods now look at the duplicates and
7.3 SST analysis and snow-ice analysis (Regular latitude/longitude grid)

Sea surface temperature analysis 1 degree by 1 degree latitude-longitude grid
Snow-ice analysis 2 degree by 2 degree latitude-longitude grid
pick a "best" report. They do not combine reports. This avoids the rather rare chance of mixing together apples and oranges because the report ID or time is wrong.

There are some problems when data from duplicates are not merged. The data from many national archives only has data at mandatory levels for raobs. To make analyses we want the significant levels whenever possible, and we want any reports that give winds by height (or pressure). Some of the data is more likely to be in the GTS data than in the national archives. This is an issue that should be thought about.

7. Data for tropical centers

NCAR has a dataset of tropical storm centers prepared by Charles Neumann. It is probably the world's best set. NMC is not using it for recent years. The forecast is often enough wrong that if the storm center data is used, then pairs of storms are often produced in the analysis. The question is whether this data should be used in earlier years.

8. The time of observations

The fact that the actual time of observations is different from the standard hours (00Z, 12Z) is still a fairly small problem. However, it is well to remember that rawinsonde balloons are usually released 1 hour before prime time, and the flight takes about 2 hours. The time of a surface observation is usually about 10 minutes before the prime hour (00Z, 06Z, etc.).

9. New time for rawinsondes (in 1957)

In the early 1950s, rawinsondes were taken at 03Z and 15Z. On 1 June 1957, the primary times were changed to 00Z and 12Z, and they are still taken at these times. A few things will need to be changed to analyze the early years.

10. Surface observations

The COADS project is now improving the world surface ocean data for 1947-79

- Finally, NMC will have COADS for 1947-93.

NCAR is preparing several big datasets of surface land observations. NMC has received data for recent years.

- NMC will finally have land synop data for 1948-94. A text is available.
11. Aircraft data

NMC has data from NCAR starting 1985. NMC will finally receive NCAR data from the NMC decode of GTS for 1962-94. Other data have been added. Some data for earlier years are also available.
Status of NMC/NCAR Reanalysis, Dec 1994

This text shows the progress in producing reanalyses at NMC. It also describes how much of the output data has been sent to NCAR.

1. Summary of NMC/NCAR reanalysis status on different dates

The progress of reanalysis is listed here and is shown in Figure 1.

a. 1992-94: Lots of work to get data and assimilation methods in place.

b. Mid-June 1994: Production of reanalyses starts at NMC.


d. July 14: Analyses are up through 4 March 1986. Earlier, NMC found that analyses for Jan 1986 were bad. The reason was that TOVS satellite soundings for the first 1 or 2 days of Jan were bad. So NMC did Jan 1986 again.

e. About July 19: Reanalysis has been completed up to July 1986 (18 months). By late July, 24 months were done.

f. 1 Aug 1994: Great Plains too warm at the surface in summer 1985. This affected some other world areas too. Fix surface treatment. Started reanalysis again about Aug 10.

h. 26 Sep 1994: 2 years were completed (1985 and 1986); about to start 1987.

i. 4 Nov 1994: Reanalyses is done for Jan 1985 to 14 Mar 1988 (now), about 3.3 years.
   - Side reanalysis test not using SSM/I ocean winds is done for Jul-Dec 1987. Jan-Jun 1988 will also be done.

j. 17 Nov 1994: As of this morning, all of 1988 is done (4 years total). It will take 2 days to get dupes out of satellite sounder data in 1989. Then reanalysis for 1989 will start.

k. 28 Nov 1994: Mar 1989 is almost done (4.3 years total). There have been some machine troubles.

l. 7 Dec 1994: NMC is now doing reanalysis for mid Oct 1989.
Figure 1. Status of reanalysis on 17 Nov 1994 (after 5.1 months of production). Analysis of data for 1985-89.

2. **NCAR sends observed data to NMC in yearly batches**

   NMC prefers to obtain data in yearly batches. NCAR has been working on many projects to prepare the data for all types (rawinsondes, winds, aircraft, satellite cloud winds, satellite soundings, surface land, and surface marine). These are for the time period 1948-93. Various documents are available that describe the data and the data coverage. This has been a huge project. The status of the data flow is:

   - NMC has received data for 1985-91.
   - NCAR is ready to send data for 1992 and 1993, after filling in raobs missing on the NMC tapes for 18 months, blocks 97 and 98. ECMWF has now also sent gap fillers to NCAR.
   - Then NCAR will send older years of data to NMC. The plan is to keep sending data in annual batches, and to stay a few years ahead of where NMC is working, in operations, on the reanalysis.

3. **Resolution of the new analyses**

   The analyses are prepared by analysis methods (and associated forecast models) that are T62 resolution, and have 28 levels in the sigma coordinate system. The sigma coordinate data includes a reasonable boundary layer; data in pressure coordinates does not. The primary analysis fields are described by spectral coefficients. Most users will use the equivalent grid point data, either on the Gaussian grid or on a 2.5° global lat-lon grid.
The horizontal resolution of the T62 Gaussian grid is 209 km (1.88°). The global grid has 192*94 points. The resolution of a 2.5° grid is 278 km at the equator. A global 2.5° grid has about 145*73 points.

4. How many years of reanalysis? The schedule?

The goal is to do at least 1957-94 (38 years), probably at least 40 years, and perhaps we will do earlier years also. We have seen that 5 years of data (1985-89) will be done by mid Dec 1994. The first task is to do years 1990-94. The year 1994 should be finished by the end of Mar 1995. By the end of 1995, a total of about 20 years will be completed.

5. Providing the reanalysis output data to NCAR (Nov 1994)

On Oct 26, NMC was about ready to ship tapes with data for the first year of reanalysis (1985). These were sent by surface mail. NCAR received the data for 1985 on 7 Nov 1994 (144 main tapes and 6 blank tapes). The main tapes have about 420 MB each.

- NCAR received all regular data for 1985 on 7 Nov 1994. The total volume of data on the 144 tapes for 1985 was 51.059693 Gbytes. Some more data will come (forecasts, etc.).

- On Nov 17, the tapes for 1986 have been copied and are ready to ship to NCAR.
  - The tapes arrived at NCAR on Nov 28 (the postmark was 5 days earlier).

- 28 Nov 1994: Data for the third year (1987) is almost ready to send to NCAR. One tape is bad and has to be fixed.

- 5 Dec 1994 (Monday): Data for 1988 was copied over the weekend at NMC, but one tape has to be fixed.

6. Description of the files of data from reanalysis

The output files from reanalysis are shown in Table 1 and described in Attachment 1. The descriptions are still not complete, but there is enough information to give users quite a good idea of what data are output, and the data volume.

7. How will NCAR make the data available

- Put the data on the mass store at NCAR.

There are many people who use the NCAR computers. Any of these people can access our data from the mass store if they want to. During last year (FY 94), there were the following number of separate users of the main equipment at NCAR:
  - 566 from NCAR or UCAR groups
  - 755 from universities
  - 116 in "other" category (Cray grants, sales, etc.)
  (total of 1437 users)
• Put the data on Exabyte tapes at a big cost reduction for users. The user will get a package deal, like a CD-ROM.

• Make the data available on the normal variety of tapes, at normal costs.

• There will be a paper about methods and costs to distribute data. Part of this information is in the Aug 1994 report (Status of Reanalysis).

• Also, NMC is putting a selection of data on a CD-ROM.

8. Forecast data from reanalysis (17 Nov 1994)

An 8-day forecast is run each 5 days. These have been run for all 4 years, 1985-88 (as of 17 Nov 1994). The data files have the whole stack of 2.5° pressure grids, each 12 hours, in the forecast; they also have the flux fields each 12 hours, but not quite as many variables. The data units are often average rates (as for precip) in each 12-hour period. The starting (time 0) grids are not in the output files. The data requires about 7 cartridge tapes each year. A year/month is together, but fields are like time series within a month. NMC will soon send all 4 years of 8-day forecast data to NCAR. To prepare the forecasts, NMC starts with one of the pairs of data files (each 5 days) called sanl and sfcnl.

9. Formats of data from reanalysis

The data for most users is in GRIB format (the grid fields), or in BUFR (the observations). Some of the most basic analysis products are in 32-bit binary (IBM floating point) GRIB and BUFR are recognized international formats by WMO.

The weather and ocean operational centers of the world are now either using these formats or going to use them. This includes the main centers I know (NMC, ECMWF, Navy-Monterey, etc.).

How do these formats compare with formats like HDF (by the University of Illinois, used so far by NASA on EOSDIS), NetCDF (by UCAR), or CDF (the parent format by Goddard)? GRIB and BUFR give more efficient packing. They are complicated formats, like the others, but they have the property that we can give the user simple software to use to extract grids of data that are unpacked from the format. This is an important property. For the other formats, the user must load a lot of software, and they must use a thick manual to understand how to retrieve the data.


Bob Kistler, NMC, says that the reanalysis is now up to 14 Mar 1988. The main stream of reanalysis is using ocean surface winds derived from SSM/I. The SSM/I data started July 1987. NMC has also reanalyzed the 6-month period (Jul-Dec 1987), not using SSM/I winds. A complete archive of this test run for comparisons is not saved. Kistler will also do Jan-Jun 1988 without SSM/I winds when time permits.
Precipitable water from SSM/I is in the dataset (starts July 1987) but is not used in the analyses.

11. The output data for a year

The files of output data are described in Attachment 1. Attachments 2, 3, and 4 present lists of variables in the different data files.

12. Users of a few primary types of data

The total output from reanalysis is over 50 Gbytes per year (same as about 400 high-density round tapes). This is a lot of data. A full 40 years will be about 2 terabytes of data. However, most people do not have to deal with all this data. Consider a few popular types of data with the annual volume:

a. grbsan1 Analyses on sigma levels, on a Gaussian grid (4909 MB/year)
b. pgb.f00 Analysis stack, 2.5° pressure (2533 MB)
c. prs Analyses, 2.5° pressure time sort (2152 MB)
d. grb2d Synop sort fluxes, Gaussian grid (1989 MB)
e. 2D Time sort of fluxes, plus other (2606 MB)

13. COADS update for 1992-93

In July 1993, NCAR sent NMC a copy of the major COADS 1980s update with data for 1980-92 (from the COADS project by NCAR, ERL, NCDC). These data have been used by Dick Reynolds to prepare weekly -1° SST data, and they are used for reanalysis. More ship data have become available. About 12 Dec 1994, a new tape (or on Internet) of all COADS data for 1992 and 1993 will be sent to NMC. This complete update should replace previous data for 1992. It will include delayed data from four main countries for 1992 (including France), and three for 1993 (U.S., U.K. Germany).

14. NMC has completed 2 years of reanalysis (written 26 Sep 1994)

Two weeks ago Bob Kistler at NMC was overwhelmed with lots of troubles: (1) some files had been lost by the storage management system, and he had to get them back; (2) the Y-YP had a few cases of hardware trouble; (3) keeping reanalysis going; and (4) how to hand out data. Today I asked if he was getting out of the pit. He said he was crawling over the rim.

NMC has completed reanalyses for 1985 and 1986. They are about to start 1987. They are still doing lots of preprocessing on the data for 1987. They look at data counts and do many other checks.

Kistler is making an internal backup for the data, and it will be in the form of what we get at NCAR. The code is 90% ready to make the backup. They are waiting for blank tapes to deliver data. He wants to use 2x density tapes. Probably only one copy of the tapes will be sent to the two groups in Boulder (this is true). NCAR will copy the tapes and send them on to Blackmon’s group.

- 5 -
15. Lose the NMC computer for reanalysis? No (written 26 Sep 1994)

NMC has been using a Y-MP8 for reanalysis, plus a small Cray EL for tape management. On Sep 17 I saw Joe Friday, head of NWS. He said they had a crisis about the Y-MP—no money to pay for it—but maybe they could replace it with smaller computers. The main computer at NMC is the Cray C-90 (16 processors), which they got in Dec 1993.

Kistler said things looked very bad for awhile (their computer would vanish). It appears okay now, but the Y-MP (with 8 processors) will depart. In May 1995 they will get a Cray Jedi with 16 processors. A little later they will get another Jedi (16 processors) and the Y-MP will depart. Each Jedi processor will have 0.6 of the power of a Y-MP processor; therefore, they will have more power than the Y-MP8 for less cost. They now are using the power of 8 Y-MP processors. The two Jedis will have the power of about 19.2 Y-MP processors. For comparison, the C90-16 has the power of about 34 Y-MP processors. It is used for the main production of forecasts at NMC.

A word of caution about the computing power is needed. The Y-MP has a solid state disk (SSD). The SSD helps a lot with rapid data manipulation; it will not be available on the Jedis. Some methods may have to be changed to keep the data processing fast.

16. NMC status reports about reanalysis

NMC has an internal meeting about reanalysis each month and prepares notes from the meeting. Examples of recent meeting dates and memo dates are:

<table>
<thead>
<tr>
<th>Meeting Date</th>
<th>Date of Memo</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 Nov 1994</td>
<td>23 Nov 1994</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* The memo dated Oct 26, 1994, has a chart which shows that the reanalysis for 1985 fits the earth’s angular momentum (observations) better than the original NMC analyses.

17. Texts at NCAR about reanalysis

* Status of Reanalysis (Aug 1994)

This text gives the status of the NMC/NCAR project and a description and data volume of the output files. A little information about the ECMWF project is included.

* A text "Data for Reanalysis; Inventories" (written Nov 1992) shows the coverage of data for about 1948 to recent years. One should also refer to later texts.

* Various texts about different types of data; ask for a list.
Table 1. Output Data from Reanalysis
This table gives the name of the data files, the volume for 1985, and the type of data. See the text for more details.

<table>
<thead>
<tr>
<th>No.</th>
<th>File</th>
<th>Mbytes (1985)</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>grbsanl</td>
<td>4909</td>
<td>The GRIB analyses for users (most don’t use guess)</td>
</tr>
<tr>
<td>1.2</td>
<td>grbsf06</td>
<td>4895</td>
<td>Sigma analyses, Gaussian grid</td>
</tr>
<tr>
<td>1.3</td>
<td>pgb.f00</td>
<td>2533</td>
<td>Analysis stack, 2.5° pressure</td>
</tr>
<tr>
<td>1.4</td>
<td>pgb.f06</td>
<td>2530</td>
<td>Guess stack, 2.5° pressure</td>
</tr>
<tr>
<td>1.5</td>
<td>prs</td>
<td>2152</td>
<td>Time sort analyses, 2.5° pressure</td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>grb2d</td>
<td>1989</td>
<td>The flux files (GRIB format) for users</td>
</tr>
<tr>
<td>2.2</td>
<td>2D</td>
<td>2606</td>
<td>Synop sort fluxes, Gaussian grid, GRIB format</td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>ipvanl</td>
<td>1462</td>
<td>Analyses on theta derived from sanl (synop sort)</td>
</tr>
<tr>
<td>3.2</td>
<td>theta</td>
<td>1971</td>
<td>Data on for 11 theta surfaces, 11 variables each (time)</td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>grib</td>
<td>269</td>
<td>Means and statistics</td>
</tr>
<tr>
<td>4.2</td>
<td>znl.f00</td>
<td>52</td>
<td>Monthly means, variance of pgb.f00, grb2d and ipvanl</td>
</tr>
<tr>
<td>4.3</td>
<td>znl.f06</td>
<td>52</td>
<td>Zonal band averages of analyses (each 6 hours)</td>
</tr>
<tr>
<td>4.4</td>
<td>optavg</td>
<td>30</td>
<td>Zonal band averages of guess (each 6 hours)</td>
</tr>
<tr>
<td>4.5</td>
<td>3D</td>
<td>192</td>
<td>Optimal averages (each 6 hours)</td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td>e 3618</td>
<td>3D heating, etc., monthly statistics from grb3d</td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td>7276</td>
<td>Data for CD-ROM (6-hour data and statistics)</td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td>2697</td>
<td>Forecast data. An 8-day forecast each 5 days</td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>sanl</td>
<td>1692</td>
<td>All basic analysis data (sigma), spectral</td>
</tr>
<tr>
<td>8.2</td>
<td>sfcsl</td>
<td>1549</td>
<td>Has all flux fields, etc., from 6-hour forecast</td>
</tr>
<tr>
<td>8.3</td>
<td>sges</td>
<td>2799</td>
<td>Like sanl, only from end of 6-hour forecast</td>
</tr>
<tr>
<td>8.4</td>
<td>bges</td>
<td>9381</td>
<td>Four basic (32-bit binary) files follow</td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>grb3d</td>
<td>2533</td>
<td>Model diagnostics each 6 hours on sigma levels, GRIB format, Gaussian grid</td>
</tr>
</tbody>
</table>
Attachment 1

The Reanalysis Output Data

Jenne has a text "Status of Reanalysis," Aug 1994, which has a lot of information not in this text. It contains lists of variables and levels given in different output files. There have been some changes with time. The metadata will gradually be clarified.

Some information about various data files

We will now provide more information about some of the data files and include more of the parent-child relationships.

1. The dataset of observations (filename is prepqm)

NCAR is supplying NMC with most of the necessary observations (raobs, surface synop, marine synop [COADS], aircraft, cloud winds, satellite sounders, etc.). NMC adds some more from other sources. NMC makes a file of all these data (prepqm). The data includes information by NMC to show how the observations were handled in the reanalysis. This file of NMC output is in the BUFR format.

2. Four primary datasets (these are mainly for backup, not for users)

There are four primary datasets from reanalysis that have all the input to forecasts and output from the 6-hour guess forecasts. The inputs are the basic analyses (sigma coordinates) plus the flux fields and other information needed. The outputs are the guess fields. NMC thinks of these files as mostly backups for NMC, and that users will use various derived datasets in the GRIB format. The binary files do not all have proper ID markers on each field; the GRIB files have proper IDS. I think that a few sophisticated users may want to use these binary datasets. This needs discussion.

- sanl All basic sigma analysis data, on 28 sigma levels
- sfcn1 The fluxes and other fields for a forecast run
- sges Basic 6-hour forecast guess, 28 sigma levels
- bges All fluxes from 6-hour guess

3. The next four items

The next four items give more information about the basic four files listed above and some of the files derived from them.
4. sanl:

All the basic analysis data (sigma) in 32-bit numbers. These are spectral coefficients. This is the direct output from the SSI analysis scheme at NMC. The model starts with this file (and sfcanl) to make a 6-hour forecast for a guess (or to make a longer forecast). The volume for 1985 is 2697 MB.

- grbsanl is derived from sanl. This is still in sigma coordinates and includes more variables, such as winds derived from the basic data in Sanl. This is on a Gaussian grid, in the GRIB format (4909 MB).

- pgb.f00: The analyses on standard pressure levels (2.5° grid) are derived from sanl. This is in the GRIB format (2533 MB). This file also has a few other things not on pressure levels (e.g., tropopause temperature, boundary layer data, etc.).

- prs: This is a time series version of the 2.5° analyses (2152 MB). prs means "pressure."

4.1 Analyses on isentropic surfaces

- ipvanl: Analyses on isentropic (theta) surfaces are also derived from sanl. This is in the GRIB format (1462 MB). In synoptic sort.

- theta: Also isentropic data from sanl; 11 theta surfaces and 11 variables; GRIB format (1971 MB). In time series sort.
  — other than sort order, what are the differences between ipvanl and theta?

5. sfcanl:

This file has all of the other fields that are necessary (besides sanl) to go into a forecast run. It includes SST, flux fields, etc. (32-bit numbers; 1692 MB).

- See the item below called "the flux files."

6. sges:

Like Sanl, only these are from the end of the 6-hour guess forecast (2799 MB).

- grbsf06, the sigma guess (in GRIB), Gaussian grid, is from this (4895 MB).
- pgbf06, the guess, in 2.5° pressure coordinates, is from this (2530 MB).

7. bges:

This file has all the flux fields, etc., from the 6-hour guess forecast.
8. **Data files for users from NMC/NCAR reanalysis**

We will next discuss the main data files for users. The analyses are packed into the GRIB format. The data precision desired for each field is specified and the GRIB packer routine computes how many bits per number will be needed for each specific grid.

9. **Primary data in sigma coordinates**

The most basic output data is the atmospheric data in sigma coordinates, that includes levels in the earth's boundary layer, and these levels follow a smoothed terrain. Also the flux fields on the Gaussian grid are part of the basic output.

- Summary of sigma coordinate archives and flux fields
  - (sanl) spectral T62, 28 sigma levels, each 6 hours (32-bit binary numbers). The volume is 2697 MB per year.
  - (grbsanl) (GRIB analysis). The volume is 4909 MB per year. This file is on the Gaussian grid, in GRIB format. There are the basic divergence and vorticity variables as well as derived wind components. Users will use this one.

*Note:* The lower sigma levels describe the world's boundary layer, and these levels will be important inputs for various problems. In the 28-level model, the lowest few sigma levels are at 997.50, 992.16, 985.44, 976.27, 963.84, 947.13, 924.91, 895.78, and 858.33.

10. **Data in pressure coordinates**

- (pgb.f00) synoptic order, 17 levels, 6 hourly. The volume is 2533 MB per year.
  - 17 levels, 1000 to 10 mb
  - Data each 6 hours, 2.5° grid
- (prs) time series of the 2.5° data is also available. The volume is 2152 MB per year.
- Monthly means and other monthly statistics

11. **The first guess fields**

The first guess fields (from forecast models) that are used in the analyses will also be archived. We do not expect very many people to use these data:

- (grbsf06) guess in sigma coordinate (each 6 hours, 28 levels). The volume is 4895 MB per year.
- (pgb.f06) guess fields transformed to pressure coordinate. The volume is 2530 MB per year.

12. **The flux files**

Compare the files grb2d and 2D.

grb2d has flux fields. It is in synop sort, on a Gaussian grid, and in the GRIB format. The file 2D is in time sort. It (2D) has all of the fields in grb2d (Gaussian grid), plus a part of
pgb.f00 (for example: troposphere temperature, maximum winds, lifting condensation level). The fields from pgb.f00 are on a 2.5° grid, in GRIB format. Therefore, the file 2D has a mix of grids (2.5° and Gaussian).

grb2d has everything in bges plus more. The model gives more things that are also interesting to users, so these are also put into grb2d. A flux file for 12Z has fields like 2m temperature that are valid at 12Z, and it has data like 6-hour average precip and radiation from a 6-hour guess that starts at 12Z. (I thought 12Z would be the end time but it isn’t.) This will be verified.

13. Optimum averaging of grids

The world is divided into regions, and an optimum average is given for each region, each variable, each 6 hours. We think these are 32-bit numbers. The volume is 30.3 MB. Someone at NMC is now also making monthly summaries of these grids.

14. Data on 11 potential temperature surfaces (6 hour, 2.5° grid)

See a previous item.

15. Monthly means and variances

These are in a file called "grib." It has statistics for all of the data in pgb.f00, grb2d, and ipvanl.

16. znl - zonal band averages

The files znl.f00 (analyses) and znl.f06 (from 6-hour guess) have various zonal band averages each 6 hours. They are not normal zonal averages for each latitude line of east-west grid points.

17. Internal model diagnostics (grb3d)

These archives have such fields as internal model heating rates that are needed by only a few people.
• Full set (grb3d). This has internal heating rates by radiation, condensation, etc. It has moistening and momentum terms. The data are at all sigma levels every 6 hours. It is a big dataset. The volume is 9381 MB for the year 1985.
• Reduced set. Monthly summary from the full set. The volume is only 192 MB per year.
• Question: Does a new type of reduced dataset need to be invented to reduce the volume of the full set to an intermediate size?

18. Data on CD-ROMs (one CD-ROM can hold 660 MB)

• A CD-ROM will be produced for every year of analyses
• Each will have a subset of pressure and isentropic level data
• Has many average fields
• Later CD-ROMs will have long time series for selected fields of data

19. Forecasts are made each few days

An 8-day forecast is prepared each 5 days. TDL argued for an 8-day forecast (not a 5-day forecast as in early plans), to get out to the point where there is no skill. The plan is to prepare an archive of 2.5° pressure fields (17 levels), and the Gaussian flux fields every 12 hours. These would be like files pgb.f00 and grb2d in the set of analyses.

What would the data volume be? With an 8-day forecast prepared each 5 days, there would be 16 cycles of data each 5 days. This compares with four cycles each day in the archive of analyses. Therefore the archive would be 16/20, or 80% of the size of pgb.f00 and grb2d. Therefore the volume of forecast data would be:

• Volume of pressure analysis (pgb.f00): 80% of 2533 MB per year
• Volume of flux fields (grb2d): 80% of 1989 MB per year
• Sum of the above two lines is 3618 MB per year, still an estimate.
Attachment 2

Distribution of Data on Tapes

The lowest cost method to distribute large amounts of data is usually to use tapes. The cost to copy tapes is significant, but it can be controlled. Table 1 has some information about the evolution of selected tape capability and cost. The more recent IBM and Exabyte tape drives have a data compression option built into the drives. The numbers in Table 1 are without compression.

<p>| Table 1. Magnetic Tape Information for 1960-95 |</p>
<table>
<thead>
<tr>
<th>Media Holds (in MB)</th>
<th>Media Cost per GB</th>
<th>Tapes per 100 GB</th>
<th>Tape Compress</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Round half-inch tapes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960 (7-track, 556 bpi, $13)</td>
<td>12</td>
<td>$1100</td>
<td>8330</td>
<td>No</td>
</tr>
<tr>
<td>1972 (9-track, 1600 bpi, $10)</td>
<td>37</td>
<td>270</td>
<td>2700</td>
<td>No</td>
</tr>
<tr>
<td>1980-94 (9-track, 6250 bpi, $10)</td>
<td>125</td>
<td>80</td>
<td>800</td>
<td>No</td>
</tr>
<tr>
<td>2. IBM 3480, 3490 cartridges</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb 1986 (200 MB, $14)</td>
<td>174</td>
<td>81</td>
<td>575</td>
<td>No</td>
</tr>
<tr>
<td>Sep 1989 (230 MB, $5)</td>
<td>200</td>
<td>25</td>
<td>500</td>
<td>No</td>
</tr>
<tr>
<td>May 1991 (460 MB, $5)</td>
<td>410</td>
<td>12.2</td>
<td>244</td>
<td>Yes</td>
</tr>
<tr>
<td>3. Exabyte</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988 8 mm (2.2 GB, $10)</td>
<td>2100</td>
<td>4.76</td>
<td>48</td>
<td>No</td>
</tr>
<tr>
<td>1991 8 mm (5.0 GB, $9)</td>
<td>4700</td>
<td>1.92</td>
<td>21.3</td>
<td>Yes</td>
</tr>
<tr>
<td>Jul 1994 8 mm (7.0 GB, $10)</td>
<td>6700</td>
<td>1.49</td>
<td>14.9</td>
<td>Yes</td>
</tr>
<tr>
<td>Oct 1995 8 mm (20 GB $20?)</td>
<td>1900</td>
<td>1.05</td>
<td>5.3</td>
<td>Yes</td>
</tr>
<tr>
<td>4. IBM NTP tape drives</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall 1994, NTP (10 GB)</td>
<td>9700</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>5. CD-ROM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986-94</td>
<td>660</td>
<td>5</td>
<td>152</td>
<td>No</td>
</tr>
</tbody>
</table>

R: for a 2x speed CD-ROM drive

Cost of drives: The prices of drives (the best prices) are about $12,000 for cartridges; $2100 for Exabyte tapes; $200 for CD-ROMs.

* These are data rates for a relatively fast drive that moves tapes at 150 inches per second.
by:

John E. Janowiak, Wesley Ebisuzaki, Muthuvel Chelliah, Suranjana Saha, Robert Kistler, Masao Kanamitsu and Glenn White

National Meteorological Center
NOAA/National Weather Service

* Presented by Wesley at Forecast conference July 1994

Something for Everyone:

(1) Global Change/Long-term Trends

- gridded analysis by 1 system, past and future analyses
- not free of all artificial trends, changes in observing network
  Two approaches:
  (1) try to get best analysis
  (2) keep observing network uniform in time
- needs: long homogenous time series with few variables

(2) Low Frequency Variability, General Circulation Studies

- uniform/state-of-art analysis system
- longer record (NMC global analysis starts in 1978)
- more vertical resolution (17 ps levels)
- more variables (heating, omega, etc.)
- pressure, theta, sigma vertical coordinates
- some terms for angular momentum budgets
- stratosphere (28 model levels)
- 6 hour resolution
(3) Radiation Studies

- basic fluxes at the top of the atmosphere, surface
- heating rates (by SW, LW) on sigma levels
- Cloud forcing (type II)
- can be compared with satellite data
- Clear sky fluxes

(4) Numerical Modellers

- some variables that only a numerical modeller would love
- more of a diagnostic for the forecast model rather than
  an analyzed quantity, useful for further development
- 99% of the variables calculated in the AMIP experiment
  (Atmos. Model Intercomparison Project)
- AMIP-like simulation using some forecast model

(5) Others

- wind stress over the ocean (Coupled Models)
- energy/water budget at surface

ISENTROPIC COORDINATES

- new

- 11 levels
  270 K
  280 K
  290 K
  300 K
  315 K
  330 K
  350 K
  400 K
  450 K
  550 K
  650 K

- variables
  Absolute vorticity
  Potential vorticity
  Montgomery stream-function
  BruntVaissala Frequency/2
  Relative Humidity
  Temperature
  \( \omega \)
  U, V
  mass-weighted U, V (i.e U*dp/d\Theta)
  \( \Theta \) at surface

- 6 hour, 2.5 degree x 2.5 degree grid
- monthly means/variances

Pressure Coordinates

- 17 levels (new: 600 mb, 925 mb, stratosphere)
  10, 20, 30, 50, 70, 100, 150,
  200, 250, 300, 400, 500, 600,
  700, 850, 925, and 1000 mb

6 Hours Time series

- U, V
- Geopotential height
- Temperature
- Relative/Specific Humidity (to 300 mb)
- \( \omega \) (to 100 mb)

Monthly Means

- above
- divergence, velocity potential
- vorticity, stream function
- UT, VT, \( \omega T \)
- Uq', Vq', \( \omega q' \)
- UU', UV, U\( \omega' \), VV, \( \omega' \omega' \)
- virtual temperature

Zonal Cross Sections
(Monthly mean)

Pressure surfaces

Geopotential Height *
Temperature *
U, V *
dP/dt *
Relative, specific humidity *
Meridional stream function
UV
Vq'
VT

Theta surfaces

NP
potential vorticity
pressure
relative humidity
U, V
mass-weighted U, V

* variance and monthly mean
- For high-accuracy calculation, need to use data on sigma surfaces (ex. difficult to resolve boundary layer with standard pressure surfaces)

- Two versions of sigma file:
  - GRIB 2.5 degree x 2.5 degree (3.4 MB/6 hours)
  - Spectral T62, (1.8 MB/6 hours)

- Three products calculated from Sigma files
  - Moisture Budget (J. Wang)
  - Monthly mean Sigma file
  - Various Angular momentum terms (with AER)

- Variables include:
  - U, V divergence, vorticity
  - Twrt, q
  - ln(surface pressure)
  - Diabatic heating terms
    - large-scale condensation
    - deep/shallow convection
    - vertical diffusion
    - short wave
    - long wave
  - Vertical diffusion of U/V momentum

\[ \text{NMC says "No, this is really on the Gaussian grid"} \]

\[ \text{Surface Quantities} \]

- Radiative
  - surface albedo (given)
  - skin temperature
  - cloud forcing (type II)
    - long/short wave, top/surface
    - fluxes at surface/top of atmosphere
  - clear sky fluxes at sfc/top of atmosphere

- Surface Conditions
  - ice (given)
  - skin temperature/SST
  - snow depth
  - runoff
  - potential evaporation
  - soil moisture (2 levels)

\[ \text{Atmospheric Variables} \]

- surface pressure
- mean sea-level pressure
- total, convective precipitation
- precipitable water
- surface stresses
- 10 m U/V
- 2 m T, q, RH
- cloud amounts (high/low/medium)
- gravity wave drag (u/v stress)
- sensible, latent heat fluxes
Frequency > 6 hours

- need resolution between 6 hours and 1 month

- some monthly mean files are strited by hour
  ex. 00Z precipatation

- some fields (24 hour average)
  total, convective precipitation
  U, V, T, height: 200, 500, 700, 850, 925 mb
  net short/long wave flux at surface
  sensible, latent heat flux
  OLR
  surface stress
  2 m temperature
  specific humidity 925, 850, 700, 500 mb

- 00Z time series
  potential vorticity, U, V, T at 315K, 350K
  skin temperature
  soil moisture at 5, 50 cm
  geopotential height, T at 100, 70 and 50 mb

- 00Z, 12Z time series (twice daily)
  height, T, U, V at 200, 500, 850 mb
  precipitable water
  specific humidity, temperature at 2 m
  ω at 500 mb.
  U, V at 10 m
  surface pressure
  1000 mb height.

Distribution

CD-ROM:
1 CD-ROM for every year of analyses
subset of pressure data, isentropic data
( limited vertical resolution, limited num. variables)
many average fields (mean, variances, covariances)

1 CD-ROM for time series (after N years)
very limited number of variables/levels

Anonymous FTP: nic.fb4.noaa.gov
15 GB reserved for reanalysis products
keep a limited number of time series on-line (?)
keep last year of reanalysis on-line
people must ftp the data before the next years analysis
replaces the older analysis
might get a faster transfer if you use Nic at Nile

TAPE:
Sent to NCAR, NCDC (Nat. Climate Data Center), CDC
Requests for tapes have to directed to NCAR, NCDC or CDC
NMC's official policy is not to supply tapes to outside users
(not our job, no money).

Other:
Unofficial data sites

Summary

The reanalysis project at NMC has started.

The results of reanalysis can be used by a larger community than the previous system
- homogenous system
- more analyzed fields
- 6 hr resolution
- longer period

Reanalysis (past) → CDAS (current)
past and monitoring the present.

Tried to make the data easy to use (GRIB/IEEE) and accessible.

With reanalysis, studies that could not have been previously attempted are now possible.

handout proposal for CD-ROM distribution
suggestions welcome
Date: Tue, 9 Aug 94 18:26:54 EDT
From: Eugenia Kalnay <wd23ek@sun1.wwb.noaa.gov>
To: deparker@email.meto.govt.uk
Subject: GISST and Reanalysis
Cc: ao@gfdl.gov, jnp@mines.utah.edu, jenne@ncar.ucar.edu,
    ken@msmail.nccd.noaa.gov, coughlan@ogp.noaa.gov, mooney@ogp.noaa.gov,
    mcpherson@sun1.wwb.noaa.gov, wallace@atmos.washington.edu,
    err@ecmwf.co.uk, schubert@schubert.gsfc.nasa.gov,
    mcane@ldgo.columbia.edu, shukla@cola.iges.org

Thank you for your letter of 20 July. Before answering, let me give you an update on where we are:

- **Problem in simulation of plants** -

We started the Reanalysis in June and the system we designed behaved as we were hoping, we were able to do two years (1985 and 1986) at a rate of a year in less than 3 weeks. We stopped to look at the results, and they seemed generally very reasonable except for too high T and low pp over the midwest US in July and August. It turned out that we were using a climatology of plant resistance created by SIB that was far too high over the US in the summer (almost 1000, compared to a typical value of about 50). So we have tested new fields for resistance and albedo, and are getting much better results. Naturally there are other problems also (e.g., surface solar radiation flux not as good as we would want), but those we cannot fix quickly, so they will have to wait for the second reanalysis.

We have decided to restart the reanalysis in a few days, and still plan to complete the 1985-1993 period this year. Next we'll do 1979-1984, followed by 1958-1978, and hopefully 1946-1957. We should finish all 40+ years in less than 3 years. We also decided, in response to several suggestions, including the UKMO, to save one year rather than two years per CDrom, so that we can put further vertical levels and fields. We should start producing them later this year.

In view of this schedule, we won't need the GISST data before we do 1979-1984 (for the period before 1982), and therefore we don't need it until early next year. Gisst 2.2 sounds from your description better than gisst 2.1, so we would be probably very grateful if you could provide it to us.

Please pass on this message to anyone interested in the project. We really appreciate you collaboration very much!

Sincerely, Eugenia Kalnay
Chief, Development Division
National Meteorological Center
Washington DC 20233
tel: 301-763-8005
fax: 301-763-8545
email: wd23ek@sun1.wwb.noaa.gov
or e.kalnay/omnet
Variables in different files (NMC/KC&R Reanalysis Project)

All files are 4 times a day at 00, 06, 12 and 18 GMT

1. Pressure file (analysis and 6 hour forecast guess on 2.5 degree latitude-longitude grid)

Geopotential height (gpm) at 16 levels
u-wind at 16 levels (m/s)
v-wind at 16 levels (m/s)
Relative humidity at lowest 7 levels (%)
Temperature (thermodynamic not virtual) at 16 levels (K)
Pressure vertical velocity at lowest 11 levels (Pa/s)
Surface pressure (Pa)
Surface skin temperature (K)
Pressure vertical velocity at the surface (Pa/s)
Relative humidity at the surface (%)
Temperature at tropopause level (K)
Pressure at tropopause level (Pa)
u wind at tropopause level (m/s)
v wind at tropopause level (m/s)
Vertical speed shear at tropopause level (/s)
Best (4-layer) lifted index (K)
Pressure at maximum wind level (Pa)
u wind at maximum wind level (m/s)
v wind at maximum wind level (m/s)
Surface geopotential height (gpm)
u wind at the lowest model level (m/s)
v wind at the lowest model level (m/s)
Potential temperature at the lowest model level (K)
Layer mean relative humidity (sig=1-0.40, 0.93-0.66, 0.66-0.40) (%)
Precipitable water (kg/m^2)

Standard pressure levels: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30
20,10 hPa

All parameters are instantaneous values

Note: This list was from Masao Kanamitsu (NMC) on 26 Aug 1994

Roy Janna, NCAR
2. Surface flux file (On the model Gaussian grid 192 x 64)

Zonal wind stress (N/m²) (A)
Meridional wind stress (N/m²) (A)
Sensible heat flux (W/m²) (A)
Latent heat flux (W/m²) (A)
Skin temperature (K) (I)
Soil moisture content (kg/m³) (I)
Water equivalent of accumulated snow depth (kg/m³) (I)
Downward short wave radiation flux at the top (W/m²) (A)
Upward short wave radiation flux at the top (W/m²) (A)
Upward long wave radiation flux at the top (W/m²) (A)
Downward long wave radiation flux at the surface (W/m²) (A)
Upward long wave radiation flux at the surface (W/m²) (A)
Downward short wave radiation flux at the surface (W/m²) (A)
Upward short wave radiation flux at the surface (W/m²) (A)
Total high level cloud cover (%) (A)
Nearby model level of high cloud top (int) (A)
Nearby model level of high cloud bottom (int) (A)
High cloud top temperature (K) (A)
Total middle level cloud cover (%) (A)
Nearby model level of middle level cloud top (int) (A)
Nearby model level of middle level cloud bottom (int) (A)
Middle cloud top temperature (K) (A)
Total low level cloud cover (%) (A)
Nearby model level of low cloud top (int) (A)
Nearby model level of low cloud bottom (int) (A)
Low cloud top temperature (K) (A)
6 hr total precipitation (kg/m³) (Abc)
6 hr convective precipitation (kg/m³) (Acc)
Ground heat flux (W/m²) (A)
Land-sea mask (1=land; 0=sea; 2=sea-ice) (I)
u wind at 10 m (m/s) (I)
v wind at 10 m (m/s) (I)
Temperature at 2 m (K) (I)
Specific humidity at the lowest model level (kg/kg) (I)
Surface pressure (Pa) (I)

(A) .. Average in 6 hr forecast
(I) .. Instantaneous value at forecast hour 6
(Acc) .. Accumulation in 6 hour forecast
3. Diagnostic file at FT=6 hr (on the model Gaussian grid 192 x 94)

Surface skin temperature (K) (I)
Soil moisture content (kg/m²) (I)
Water equivalent of accumulated snow depth (kg/m²) (I)
Soil temperature at 10, 50 and 500 cm (K) (I)
Surface roughness (m) (I)
Total 6 hour accumulated precipitation (kg/m²) (Abc)
Convective 6 hr accumulated precipitation (kg/m²) (Abc)
Sensible heat flux (W/m²) (A)
Zonal wind stress (N/m²) (A)
Meridional wind stress (N/m²) (A)
Latent heat flux (W/m²) (A)
Albedo (%) (I) *
Downward short wave radiation flux at the top (W/m²) (A)
Upward short wave radiation flux at the top (W/m²) (A)
Upward long wave radiation flux at the top (W/m²) (A)
Downward long wave radiation flux at the surface (W/m²) (A)
Upward long wave radiation flux at the surface (W/m²) (A)
Downward short wave radiation flux at the surface (W/m²) (A)
Upward short wave radiation flux at the surface (W/m²) (A)
Total high level cloud cover (%) (A)
Nearby model level of high cloud top (int) (A)
Nearby model level of high cloud bottom (int) (A)
Total middle level cloud cover (%) (A)
Nearby model level of middle level cloud top (int) (A)
Nearby model level of middle level cloud bottom (int) (A)
Total low level cloud cover (%) (A)
Nearby model level of low cloud top (int) (A)
Nearby model level of low cloud bottom (int) (A)
Convective cloud cover (%) (A)
Clear sky upward long wave flux (W/m²*2)
Clear sky upward solar flux (W/m²*2)
Clear sky downward long wave flux (W/m²*2)
Clear sky downward solar flux (W/m²*2)
Clear sky upward solar flux (W/m²*2)
Cloud forcing net solar flux (W/m²*2)
Cloud forcing net solar flux at the surface (W/m²*2)
Cloud forcing net solar flux (W/m²*2)
Cloud forcing net long wave flux (W/m²*2)
Cloud forcing net long wave flux (W/m²*2)
Cloud forcing net long wave flux (W/m²*2)
Visible beam downward solar flux at the surface (W/m²) (A)
Visible diffuse downward solar flux at the surface (W/m²) (A)
Near IR diffuse downward solar flux at the surface (W/m²) (A)
Large scale condensation heating rate at 28 model levels (K/s) (A)
Deep convective heating rate at 28 model levels (K/s) (A)
Deep convective moistening rate at 28 model levels (Kg/Kg/s)
Shallow convective heating rate at 28 model levels (K/s)
Shallow convective moistening rate at 28 model levels (Kg/Kg/s)
Vertical diffusion heating rate at 28 model levels (K/s)
Vertical diffusion zonal acceleration at 28 model levels (m/s^2)
Vertical diffusion meridional acceleration at 28 model levels (m/s^2)
Vertical diffusion moistening rate at 28 model levels (Kg/Kg/s)
Solar radiative heating rate at 28 model levels (K/s)
Long wave radiative heating rate at 28 model levels (K/s)

Note that some parameters are redundant with surface flux file

4. Diagnostic file at FT=0 (on model Gaussian grid 192 x 94)

Surface skin temperature (K) (I)
Soil moisture content (kg/m^2) (I)
Water equivalent of accumulated snow depth (kg/m^2) (I)
Soil temperature at 10, 50 and 500 cm (K) (I)
Surface roughness (m) (I) *

(I) Instantaneous values
*) Not a predicted parameter but changes because of the time interpolation of monthly climatology

5. Grib sigma file (on model Gaussian grid 192 x 94)

Relative vorticity at 28 model levels (1/s)
Divergence at 28 model levels (1/s)
Thermodynamic temperature (not virtual) at 28 model levels (K)
Specific humidity at 28 model levels (kg/kg)
x-gradient of log surface pressure (1/m)
y-gradient of log surface pressure (1/m)
u wind (m/s) at 28 model levels
v wind (m/s) at 28 model levels
Surface pressure (Pa)
Surface geopotential height (gpm)
x-gradient of height (m/m)
y-gradient of height (m/m)

6. Zonal diagnostic file

Multi-level fields divided into the area 90N-60N, 60N-30N, 30N-30S, 30S-60S and 60S-90S.
Global averages are also computed.

- u component of winds (I)
- v component of winds (I)
- Virtual temperature (I)
- Specific humidity (I)
- Squared vorticity (I)
- Squared divergence (I)
- Pressure vertical velocity (I)
- Temperature (I)
- Relative humidity (I)
- Kinetic energy (I)
- Convective heating rate (A)
- Large scale heating rate (A)
- Shallow convection heating rate (A)
- Heating due to vertical diffusion (A)
- Convective moistening rate (A)
- Shallow convection moistening rate (A)
- Moistening due to vertical diffusion (A)
- u-acceleration by vertical diffusion (A)
- v-acceleration by vertical diffusion (A)
- Heating by short wave radiation (A)
- Heating by long wave radiation (A)

Single level field separated by surface characteristics i.e. snow covered land/sea-ice/snow-covered-sea-ice/sea over the same latitudinal belt as for the multi-level fields

- Total precipitation (Acc)
- Convective precipitation (Acc)
- Sensible heat flux (A)
- Latent heat flux (A)
- u-stress (A)
- v-stress (A)
- Area coverage of total rain (Acc)
- Area coverage of convective rain (Acc)
- Surface pressure (I)
- Skin temperature (I)
- Soil wetness (I)
- Snow depth (I)
- 10 cm deep soil temperature (I)
- 50 cm deep soil temperature (I)
- 500 cm deep soil temperature (I) *
- Surface net short wave radiation flux (A)
- Surface net long wave radiation flux (A)
- Relative humidity at the lowest model level (I)
- Virtual temp at the lowest model level (I)
Temperature at the lowest model level (I)
Specific humidity at the lowest model level (I)
Surface roughness (I) *
Land/sea/sea-ice mask (I) *
U-acceleration by gravity wave drag (A)
V-acceleration by gravity wave drag (A)
Angular momentum budget surface torque (A)
Angular momentum budget gravity wave drag (A)
Angular momentum mountain torque (A)
Total angular momentum (I)
Planetary angular momentum (I)

(A) .. Average in 6 hr forecast
(I) .. Instantaneous value at forecast hour 6
(Acc) .. Accumulation in 6 hour forecast
*) Not a predicted parameter but changes because of the time interpolation of the monthly
climatology and snow cover change

7. Restart file (no grib packing is done to insure reproducibility).

7.1 Sigma file (T62 L28) in spectral coefficient form

Vorticity at 28 model levels (1/s)
Divergence at 28 model levels (1/s)
Virtual temperature at 28 model levels (K)
Specific humidity at 28 model levels (Kgm/Kgm)
Log of surface pressure
Surface geopotential

7.2 Surface file (Gaussian grid 192 x 94)

Surface temperature
Soil moisture
Snow depth water equivalent
10 cm deep soil temperature
50 cm deep soil temperature
500 cm deep soil temperature
Surface roughness-
Convective cloud cover
Convective cloud bottom height
Convective cloud top height
Land/sea/ice mask
Plant evaporation resistance
Ratio between 10 m winds and lowest model level winds
7.3 SST analysis and snow-ice analysis (Regular latitude/longitude grid)

Sea surface temperature analysis 1 degree by 1 degree latitude-longitude grid
Snow-ice analysis 2 degree by 2 degree latitude-longitude grid
NMC/NCAR Reanalysis Information (January 1995)

1. Status of Reanalysis on 4 Jan 1995

The new analysis of Aug-Dec 1989, using the combined NMC/ECMWF raob data, has been completed. Therefore, 5 years of reanalysis are now done (1985-89). In one or two days, NMC will start doing reanalysis for 1990.

There is now more observed data to use, so Bob Kistler at NMC says it is taking an hour of wall clock time per day on the computer, compared with 45 minutes per day for earlier years (the mid 1980s).

2. Drop in the count of raob significant wind reports

The count of significant level winds on NMC tapes was low from Aug 1989 to Sep 1991. Other sources have been used to restore the data. The reported raob wind data on GTS comes in several different transmissions. There are reports of winds and temperature at mandatory levels, plus reports of significant level temperature where changes occur, plus significant level winds, plus winds by height (about each 300m or 1000 ft in the lower levels).

In Dec 1994, it was discovered that there was a 2-year period in 1989-91 where the count of significant level winds dropped markedly on the NMC tapes. No one knows why this happened. The exact period (about 26 months) is as follows:

- The first lower count of significant wind levels is on 12Z, 1 Aug 1989. The count went from about 4600 levels for a prime observation time to 1400 or 1500 levels.

- The first time when the count went back up was 00Z, 26 Sep 1991. The count went up from about 1500 to 6500 levels on this date. NCAR has a copy of the Navy GTS tapes. These tapes could have helped a lot to restore the NMC data, but when the Navy and NMC sig wind counts were both good, the Navy counts were then only about 70 to 80% of the good NMC counts. ECMWF supplied their raobs to NMC. The combined set of NMC and ECMWF raob data is being used for the analyses.

- Since the other raob wind data was present, the lack of this transmission probably did not hurt the analyses very much, but the decision was made to do the analyses for Aug-Dec 1989 again, using the complete dataset, as described below.

ECMWF sent all of their global raobs to NMC for all days in the 26-month period, Aug 1989 through Sep 1991. NMC finds that this provided the missing data. NMC found that in addition to the sig wind problem, the ECMWF tapes had raob data that NMC did not have and NMC had data that ECMWF did not have. For example, NMC had extra data for N. America and China that ECMWF did not have. NMC will prepare information about the comparison. NMC is combining the NMC and ECMWF global raob data for this 26-month period, and the combined set is being used for the reanalysis.
Summary of NMC/NCAR Reanalysis Data Archive at NCAR

Chi-Fan Shih
<chifan@ncar.ucar.edu, 303/497-1833, fax:303/497-1298>

*This is NCAR ds090.0 dataset. All 1985 to 1988 data is now available to NCAR users. There are 276 mass storage volumes per year. Each volume is one UNIX tar file. Complete lists of data files on each volume is available in our anonymous ftp area on ncardata.ucar.edu in ftp/datasets/ds090.0 directory.

*The file sizes listed below are computed from the actual tar lists of all vsn's. The sizes of 1985 have been verified against that from NMC.

*The following files are missing in 1985:
  znl.f00  - July, August, September, October
  znl.f06  - July, August, September, October
  bges    - August

*The following individual files are missing in 1985
  optavg  - No 00Z file of the first day of months from Feb to Oct.
  grbsf06  - No grbsf0685083000, grbsf0685083006 two files.

*Starting from May 1988, one new variable 'LAND.grib.yymm' is added.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>starting date</td>
<td>11/08/94</td>
<td>11/28/94</td>
<td>12/16/94</td>
<td>12/24/94</td>
</tr>
<tr>
<td>finishing date</td>
<td>11/17/94</td>
<td>12/07/94</td>
<td>12/23/94</td>
<td>01/04/94</td>
</tr>
<tr>
<td>total number of vsns</td>
<td>276</td>
<td>276</td>
<td>276</td>
<td>276</td>
</tr>
<tr>
<td>total number of files</td>
<td>32973</td>
<td>34092</td>
<td>34092</td>
<td>34155</td>
</tr>
<tr>
<td>total bytes in binary (Gb)</td>
<td>51.034</td>
<td>51.294</td>
<td>51.508</td>
<td>54.558</td>
</tr>
<tr>
<td>total bytes in cos-blocked (Gb)</td>
<td>51.05</td>
<td>51.05</td>
<td>51.05</td>
<td>51.05</td>
</tr>
<tr>
<td>GAUS used to read back</td>
<td>2.33688</td>
<td>2.08447</td>
<td>1.96962</td>
<td>1.81607</td>
</tr>
</tbody>
</table>

Files of reanalysis data now on-line

<table>
<thead>
<tr>
<th>Data</th>
<th>1985(mb)</th>
<th>1986(mb)</th>
<th>1987(mb)</th>
<th>1988(mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>grb3d</td>
<td>(GRIB)</td>
<td>9380.76</td>
<td>9405.76</td>
<td>9424.35</td>
</tr>
<tr>
<td>prepqm</td>
<td>(not GRIB)</td>
<td>7275.79</td>
<td>7241.43</td>
<td>7463.67</td>
</tr>
<tr>
<td>sanl</td>
<td>(not GRIB)</td>
<td>2697.01</td>
<td>2697.01</td>
<td>2697.01</td>
</tr>
<tr>
<td>sfcanl</td>
<td>(not GRIB)</td>
<td>1692.37</td>
<td>1692.37</td>
<td>1692.37</td>
</tr>
<tr>
<td>sges</td>
<td></td>
<td>2798.72</td>
<td>2798.72</td>
<td>2798.72</td>
</tr>
<tr>
<td>bges</td>
<td>(not GRIB)</td>
<td>1548.64</td>
<td>1692.37</td>
<td>1692.37</td>
</tr>
<tr>
<td>znl.f00</td>
<td>(not GRIB)</td>
<td>51.54</td>
<td>77.74</td>
<td>77.74</td>
</tr>
<tr>
<td>znl.f06</td>
<td></td>
<td>51.54</td>
<td>77.74</td>
<td>77.74</td>
</tr>
<tr>
<td>grbsf06</td>
<td></td>
<td>4894.62</td>
<td>4919.57</td>
<td>4920.15</td>
</tr>
<tr>
<td>grbsanl</td>
<td></td>
<td>4909.00</td>
<td>4927.23</td>
<td>4925.91</td>
</tr>
<tr>
<td>pgb.f00</td>
<td></td>
<td>2533.17</td>
<td>2540.69</td>
<td>2536.04</td>
</tr>
<tr>
<td>pgb.f06</td>
<td></td>
<td>2530.31</td>
<td>2537.57</td>
<td>2533.57</td>
</tr>
<tr>
<td>ipvanl</td>
<td>(GRIB)</td>
<td>1462.07</td>
<td>1462.80</td>
<td>1457.95</td>
</tr>
<tr>
<td>optavg</td>
<td>(not GRIB)</td>
<td>30.25</td>
<td>31.18</td>
<td>31.72</td>
</tr>
<tr>
<td>2D</td>
<td>(GRIB)</td>
<td>2605.56</td>
<td>2607.37</td>
<td>2607.37</td>
</tr>
<tr>
<td>3D + grib + prs</td>
<td></td>
<td>2612.24</td>
<td>2620.30</td>
<td>2614.80</td>
</tr>
<tr>
<td>grib</td>
<td></td>
<td>268.65</td>
<td>283.16</td>
<td>282.77</td>
</tr>
<tr>
<td>prs</td>
<td></td>
<td>2152.03</td>
<td>2159.10</td>
<td>2154.43</td>
</tr>
<tr>
<td>3D</td>
<td></td>
<td>191.55</td>
<td>178.04</td>
<td>177.59</td>
</tr>
</tbody>
</table>

* This tape will be replaced
Data Status from NMC/NCAR Reanalysis

We will describe the status of the NMC/NCAR 40-year reanalysis project and the associated data archives. One CD-ROM per year has a good-sized sample of the data from reanalysis; the first one probably will be available about late Jun 1995. SSMI wind observations were used for the surface ocean for Jul 1987-on, but a problem caused them to be biased about 2m/sec high. This caused the Hadley cell to be too intense and global rainfall increased by about 10%. Nearly all of this problem could be removed by fixing the error, but it was finally decided that SSMI winds would not be used. Analyses for Jul 1987-on will be replaced with new calculations. Appendix A has more information about SSMI.

Reanalysis will cover the period 1957-96 and perhaps earlier

1. When Will Data be Available?

NCAR has data from reanalysis for 1985 through 1991 (7 years), but data for Jul 1987 through 1991 will be replaced because of the SSMI problem. The dates for data preparation are roughly as follows:

- Good data for Jan 1985 - Jun 1987 (2.5 years) are now at NCAR.
- Calculations are done for Jan-Jun 1982.
- Data for Jul 1987 - Dec 1987 (without SSMI) were finished on Apr 19.
- April 26 - All data for 1985-89 have been calculated at NMC. They are starting data for 1990 today. The years 1988-89 went fast because part of the "no SSMI" calculations had been made earlier. So a total of 5.5 years are done.
- By early Sep 1995 we will have 10.5 years done (1985-94, and Jan-Jun 1982).
- In Dec 1995, data for 1982-94 (13 years) will be ready.
- Then the project will start on earlier data.

10 years of output by Sep 1995; 13 years by Dec 1995

Note: The Cray Y-MP being used for reanalysis at NMC will be replaced with two very good, newer, and less expensive small Crays. This means extra software work at NMC in Jun-Jul, 1995.

Data Volume: Total output is over 50 Gbytes per month, but most users will be happy with smaller subsets of data, such as all analyses in either pressure or sigma coordinates.

2. Description of the Data Output from the NMC/NCAR Reanalysis

Table 1 summarizes the output from reanalysis. The main descriptions are in two sources:

- "Status of NMC/NCAR Reanalysis, Dec 1994" by Jenne
3. The Change in Types of Satellite Sounder Data, Oct 1978 - Feb 1979

The VTPR data were used operationally at NMC until 28 Feb 1979, then TOVS data were used. There is an overlap of data from TOVS and VTPR for several months, which will permit tests of the pre-TOVS analyses. The archives that exist are:

- TOVS basic radiances (29 Oct 1978-present)

- TOVS 2.5° soundings (1 Jan 1979-present). These have cloud-cleared radiances and soundings. Some early data were hurt by rain, so that tests on radiances are needed to determine good vs. bad data.


NCAR has tapes with the NOAA VTPR sounding data for 20 Nov 1972 to 28 Feb 1979. This has an overlap of several months with the TOVS sounders. The VTPR tapes have three types of files: the basic radiances, cloud-cleared radiances, and the derived soundings. All of these data for the whole 6.4 years have a volume of only 5.516 Gbytes.

NCAR will send a copy of the whole dataset to NMC and they will extract the portions of the data that they need to use now.

4. Some Bad Locations in TOVS 2.5° Data

There are some bad locations in the TOVS 2.5° data. In most cases these errors only happen on the first 1 or 2 days of a new year, and not all years.

NMC has done diagnostics on much of the TOVS 2.5° data and has found three cases of bad data; the first 2 days of 1986 and 1987 are bad and there was also a bad case in Dec of 1982 or 1984. We will get the date.

ECMWF is using TOVS 1b from NCAR to calculate cloud-cleared radiances and soundings for some gaps in the TOVS 2.5° data. Rex Gibson (ECMWF) sent a message on 24 Apr 1995:

"We have just produced the first locally regenerated CCR data (cloud cleared) from full NESDIS 1b data, using the system we hope to use for gaps in the 250 km CCR data. We have also, with help from NESDIS, been successful in recovering some of the mislocated satellite data by recomputing positions from orbital information."

5. Need New SST Grids before 1982

Global SST analyses are needed for surface boundary conditions for reanalysis. Dick Reynolds, NMC, has good global SST grids (weekly, 1°) for the period Nov 1981-on. These
use both COADS data and satellite data, and they properly blend with the sea ice edges that are known. They are being used for reanalysis.

Before Nov 1981, a better set of SST grids are needed for reanalysis than are now available. The UK Met office is working on grids, but a direct analysis using data that are usually sparse loses much of the Pacific equatorial SST pattern that is associated with El Niño changes. Reynolds has been working on methods to get the patterns right. Now Reynolds and the UK are working together. SST grids for the period 1957-81 will be ready by about Jul 1995, for use in reanalysis.

Monthly SST from Reynolds, 1950-92: Dick Reynolds has used the excellent SST data from Nov 1981-93 to make an EOF analysis and determine robust monthly structure functions that can be applied to older data to obtain a better pattern of SST fields than can be derived by other methods. He has produced a monthly 2° x 2° SST grid for 1950-92 using COADS data, for 66°N to 45°S. In these grids, ice edges are not used. The final analysis is from many structure functions; the data aren't applied again. The UK will also use this method. They may then introduce observations again as a final step and they will introduce ice edges and complete the global analysis. These grids will be used for reanalysis. NCAR will have a copy of the data from Reynolds.

6. Cloud Drift Winds from Satellite Data

These data are used from the original NMC tapes at NCAR. Also, JMA provided a copy of their GMS cloud winds for the period 1978-91. The JMA data were sent directly to NMC.

A message from Rex Gibson (ECMWF, Apr 1995) says that they are using the "revised GMS cloud winds from JMA" from Apr 1981-on. I hope that NMC and ECMWF have the same GMS data??

7. High-resolution Raob Data for TOGA COARE

TOGA COARE was an equatorial Pacific experiment, near Indonesia, that operated in late 1992 and early 1993. There were extra ship, island, and aircraft data. By 25 Jan 1995, the raob dataset had a total of 5,000 soundings out of 15,000 to be expected when data preparation is finished. The data are online at NCAR. See the 25 Jan 1995 memo by Jenne.

8. Data Inputs for 1994

It would have been possible to add some delayed data for 1994 to use in reanalysis. For example, some delayed ship data for much of the year could have been ready by Apr 1995. However, NMC prefers to use the data for 1994 the same as it was for operations, so NCAR just sent this data back to NMC (but in the style for reanalysis).

9. Merging of separate raobs

Jack Woollen (NMC) has some good ideas about merging raobs, and Bob Kistler (NMC) suggests that I talk with him. The merge is done at NMC.
10. New COADS ship data

The COADS project is preparing new data for 1947-79 and is processing data for the 1970s first. We anticipate that data for the 1970s will be ready to send to NMC about Sep 1995.

11. Use of Australian Surface Bogus Data

NMC observed a copy of Australian surface bogus data via ECMWF. The data are being used. The data go back to about 1972. We will check on the starting date.

12. Other Data

This text only describes a few of the data issues and sources. NCAR has a number of texts that describe the data issues and sources in much more detail.


The count of significant wind reports on the original NMC tapes was low from Aug 1989 to Sep 1991. The main reports of raob temperature and wind were still available in the data. NCAR could have filled in many of the wind reports from Navy tapes, but ECMWF data was more complete. ECMWF sent all of their global raob data for Aug 1989 through Sep 1991 to NMC. There were regions where NMC tapes had data missing from ECMWF, so NMC combined the two sources for this period. Jenne has a 1-page report, "NMC/NCAR Reanalysis Information," dated 4 Jan 1995, that describes this in more detail. This problem was handled in the last half of 1994. The data got into reanalysis.


Some lost raobs were replaced: Sometime in 1991, NMC put a limit on the size of the buffer to save raob data, and the station catalog list was also involved. As time went on, a lot of raobs were lost from many high block-station raobs (which also probably came in late). The loss gradually started about Jan 1992, and was fixed at NMC on about Jun 29, 1993. It especially affected the receipt of block 97 and 98 data. For example, about 30 good raob stations had almost no data during the first half of 1993 (and earlier) on these tapes. The problem was discovered at NCAR when people started looking for data to help the TOGA COARE experiment. NCAR found some of the missing data in the Navy source, but ECMWF had better coverage. ECMWF provided raob and piball data for blocks 97 and 98 for the period.

15. Some Pacific UA Data from Japan

JMA (Japan) compared their data receipt over China and the Pacific with ECMWF. There were some stations (raob and pibal) that JMA obtained that ECMWF did not. NMC obtained a copy of these data from Japan. Jenne has a copy (via Kanamitsu at NMC) of two plots from Japan, dated Aug 1993. These show where the extra stations are located.
The JMA raob filler data covers the period 1979-93. Jack Woollen at NMC found that most of these data are also in the other sources from NCAR, but he estimates that about 10% to 20% of it got through his duplicate checks. A better estimate could be prepared if necessary.


<table>
<thead>
<tr>
<th>Jan 1988 - Apr 1988</th>
<th>This data will be analyzed next (no SSMI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 1988 - Dec 1989</td>
<td>NMC already did &quot;no SSMI&quot; analyses for this period, but Kistler has to generate products. This will go fast.</td>
</tr>
<tr>
<td>Jan 1990 - Dec 1994</td>
<td>Still need to do a full analysis of this period, without using SSMI.</td>
</tr>
</tbody>
</table>

*Note:* Data for 1985, 1986, and 1987 were analyzed at NMC using version 1 of data preprocessing. Data for July 1987-on will be analyzed using version 2.

17. How the Surface is Handled

Reanalysis now has an impressive way to handle the surface conditions. During the testing of the first 2 years of reanalysis, it was found (in Aug 1994) that surface temperatures over part of the U.S. Great Plains got too warm by about 5°C in Jul and Aug. Some other parts of the world were also affected. The way that the growing crop vegetation was described and handled was then improved. All reanalysis output is based on the new methods. Information about this problem is included for historical reasons and to note that the surface atmosphere is quite sensitive to the vegetation.

This problem did not affect data from the operational NMC model because it was then using the older physics. The new surface methods for reanalysis went into production in the operational model in Jan 1995.

More information about the surface conditions is in:

- NMC monthly report, about Aug-Sep 1994, by Kalnay

18. Summary

After a temporary setback related to some calculations of SSMI winds, the reanalysis project is again progressing at a rapid clip. NMC has to undergo a computer conversion, but we think that 10.5 years of data will be ready in Sep 1995 and 13 years in Dec 1995.
Table 1. Output Data from Reanalysis

This table gives the name of the data files, the volume for 1985, and the type of data. See the text for more details.

<table>
<thead>
<tr>
<th>No.</th>
<th>File</th>
<th>Mbytes (1985)</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>grbsanl</td>
<td>4909</td>
<td>The GRIB analyses for users (most don't use guess)</td>
</tr>
<tr>
<td>1.1</td>
<td></td>
<td></td>
<td>Sigma analyses, Gaussian grid</td>
</tr>
<tr>
<td>1.2</td>
<td>grbsf06</td>
<td>4895</td>
<td>Sigma guess, Gaussian grid</td>
</tr>
<tr>
<td>1.3</td>
<td>pgb.f00</td>
<td>2533</td>
<td>Analysis stack, 2.5* pressure</td>
</tr>
<tr>
<td>1.4</td>
<td>pgb.f06</td>
<td>2530</td>
<td>Guess stack, 2.5* pressure</td>
</tr>
<tr>
<td>1.5</td>
<td>prs</td>
<td>2152</td>
<td>Time sort analyses, 2.5* pressure</td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td>The flux files (GRIB format) for users</td>
</tr>
<tr>
<td>2.1</td>
<td>grb2d</td>
<td>1989</td>
<td>Synop sort fluxes, Gaussian grid, GRIB format</td>
</tr>
<tr>
<td>2.2</td>
<td>2D</td>
<td>2606</td>
<td>Time sort, all of grb2d plus some 2.5* data from pgb.f00</td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td>Data for isentropic surfaces (GRIB format)</td>
</tr>
<tr>
<td>3.1</td>
<td>ipvanl</td>
<td>1462</td>
<td>Analyses on theta derived from sanl (synop sort)</td>
</tr>
<tr>
<td>3.2</td>
<td>theta</td>
<td>1971</td>
<td>Data for 11 theta surfaces, 11 variables each (time)</td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td>Means and statistics</td>
</tr>
<tr>
<td>4.1</td>
<td>grib</td>
<td>269</td>
<td>Monthly means, variance of pgb.f00, grb2d and ipvanl</td>
</tr>
<tr>
<td>4.2</td>
<td>znf.f00</td>
<td>52</td>
<td>Zonal band averages of analyses (each 6 hours)</td>
</tr>
<tr>
<td>4.3</td>
<td>znf.f06</td>
<td>52</td>
<td>Zonal band averages of guess (each 6 hours)</td>
</tr>
<tr>
<td>4.4</td>
<td>optavg</td>
<td>30</td>
<td>Optimal averages (each 6 hours)</td>
</tr>
<tr>
<td>4.5</td>
<td>3D</td>
<td>192</td>
<td>3D heating, etc., monthly statistics from grb3d</td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td>e 3618</td>
<td>Data for CD-ROM (6-hour data and statistics)</td>
</tr>
<tr>
<td>6.</td>
<td>preppqm</td>
<td>7276</td>
<td>Forecast data. An 8-day forecast each 5 days</td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td>Observed data with all metadata (BUFR format)</td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td>Four basic (32-bit binary) files follow</td>
</tr>
<tr>
<td>8.1</td>
<td>sanl</td>
<td>2697</td>
<td>All basic analysis data (sigma), spectral</td>
</tr>
<tr>
<td>8.2</td>
<td>sfcanl</td>
<td>1692</td>
<td>all other fields for forecast run (SST, flux, etc.)</td>
</tr>
<tr>
<td>8.3</td>
<td>sges</td>
<td>2799</td>
<td>Like sanl, only from end of 6-hour forecast</td>
</tr>
<tr>
<td>8.4</td>
<td>bges</td>
<td>1549</td>
<td>Has all flux fields, etc., from 6-hour forecast</td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td>3D heating, momentum, and moistening follows</td>
</tr>
<tr>
<td>9.1</td>
<td>grb3d</td>
<td>9381</td>
<td>Model diagnostics each 6 hours on sigma levels, GRIB format, Gaussian grid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total 54.65 Gbytes, one year</td>
</tr>
</tbody>
</table>

For backup and restart, not for most users.
Appendix A

The Use of SSMI Wind Data

Once we knew that the SSMI winds used in the reanalysis were biased high, the first idea was that NMC would fix the bias and still use the winds. It was found that the bias could be removed, but that some questions would remain.

A-1 Decision to Drop SSMI Wind Data from Reanalysis

SSMI is a satellite microwave instrument that permits us to calculate ocean surface wind speeds (and other variables). Tests showed a little gain from using SSMI winds for reanalysis, but not much. It would take a lot of processing to use it, and would slow the project down. And NMC would have to pay ~$30,000 for SSMI data for 1992, 1993, 1994. On Friday, Apr 7, NMC talked with Wentz by phone and he expressed some of the same worries about the algorithms being used to calculate SSMI winds that he told me about earlier. So still more tests would be needed if we used SSMI data. Frank Wentz is a key scientist on SSMI data. He also runs a small company in California.

On Apr 10, NMC decided that it was best now to leave out SSMI data and try to get the reanalysis project back on schedule. The project is now moving forward at a fast clip. It has been a trying time at NMC to sort all of this out. People are now feeling much happier.

A-2 Texts about SSMI Data

Jenne has three short texts about SSMI:

- "SSMI Data (Primary and Products)," Apr 1995, 1 pp.

Note: Wentz has calculated surface winds and other products from SSMI data. NCAR ordered all of the product data (14 satellite years) for Jul 1987 - Dec 1994. The data arrived 18 Apr 1995. The data were received on 14 Exabyte tapes. The volume is probably about 65 Gbytes. The cost was $2800.
A-3 The Use of SSMI Data in NMC Operational Analyses

NMC started using SSMI data in operations on 4 Feb 1992. The winds were calculated from basic data in satellite coordinates. The speed bias problem that had to be fixed in reanalysis did not happen in the operational product.

A-4 Message from NMC (March 10, 1995), found trouble with SSMI

We discovered that the SSMI wind speeds we have assimilated for the period July 1987 - Dec 1991 were computed without a transformation from "antenna" temperature to "brightness" temperature. A preliminary evaluation estimated that the error created 10m wind speeds that are 2 m/sec too fast. Moreover, further investigation has revealed occasional overwriting of a partial orbit from tomorrow on today's 00Z collection.

We need a period of time to assess how best to replace the data already produced. This bias is very similar to the jump that we found when the SSMI data was introduced, and which we then attributed to model bias. The good news is that with the corrected SSMI winds the reanalysis jump will be much smaller (the SSMI will still provide a realistic distribution of the surface winds).

In order to remain productive during this assessment period, NCAR is going to supply us with the data for the period 1982 -1984. We might reasonably expect to begin running that period in the second half of March, estimating to finish by early May. This period of delay will also allow the UKMO (David Parker and Nick Rayner) to work on their GISST (sea surface temperature), taking into account the EOF SST analysis ideas suggested by Reynolds et al.

While we are running this pre-SSMI period, we will take the time to decide among the following three choices:

1. recover the SSMI data we have, or
2. access the operational NESDIS archive, or
3. revert to no SSMI data

The earliest we then could expect to reach the end of 1994 would be around October. However, we are slated for a computer replacement during this same period. So realistically, it may not be until the end of the year before we complete 1982-1994.

A preliminary assessment of the impact of this wind bias error indicates that it has a very small effect on most fields, but that it increases the latent heat flux and the surface wind speed by about 10%. (We observed an increase in global precip from 2.7 to 3mm/day with the introduction of SSMI data). Although for most purposes these data would still be usable, in order to avoid confusion, we request that any distribution of the data for July 1987-end of 1991 be stopped, and if any has already occurred, that the recipients be notified that the analyses included this biased data, and that replacement data come in the time frame just mentioned.
We sincerely regret the problems that this will cause.

Eugenia Kalnay, NMC

A-5 Reanalysis was stopped in mid Feb 1995, but has been started again

(This paragraph was written Apr 7, 1995)

By mid Feb 1995, 7 years of reanalysis were done. Then reanalysis was stopped when we found that ship dew point temperatures were off a little bit in 1992 in one national ship dataset. That was fixed in a few days.

But then a worse problem was discovered. The SSMI ocean wind data was inspected more carefully. The derivations had used the wrong type of temperature, so they were biased about 2 m/sec too fast.

NCAR sent NMC data for 1981-84 so that NMC could analyze data for 1982-84, while sorting out the SSMI problems.

- NMC had planned to analyze data for Dec 1981 for spinup, and then analyze data for 1982, 1983 and 1984. But the surface does not spin up well in one month. So they fed in a December climatology, and then started with daily data for December. This worked well. Now (Apr 7) they are analyzing the data for Feb 1982.

Now production is rolling again. It is going quite a bit faster than more recent years because there is less data.
Status of NMC/NCAR Reanalysis Project

- Started production June 1994
- By Feb 1995 we had 7 years done
- In mid Feb 1995 we stopped the project
  - ship data for 1992 needed dew point fix (quick)
  - ocean surface winds from SSMI biased 2 m/sec too strong (wrong radiance temp)
  - SSMI is for Jul 1987-on
  - lost about 5 weeks to make a zillion tests
- Decided to rerun without SSMI
- On Apr 26, 5.5 years are done
- Estimate 13 years done by Dec 1995
- We are working hard to prepare older data

NCAR Scientific Computing Division
Supercomputing • Communications • Data

Dickinson led a NASA research project
Jenner gave this talk to the project PIs.

Roy Jenner
May 1995
Global Clouds and Surface Radiation (New ISCCP data)

1. ISCCP radiances (Jul 1983 to present)
   — Geosynchronous data each 3 hours
   — 8 years of B3 is about 200 Gbytes

2. Older cloud calculations (Jul 1983 to Jun 1991)

3. Making new cloud calculations
   — Started about Nov 1994
   — Did 1991 first
   — Will have 1993 about Sep 1995
   — Will have 1988 about Jan 1996
   — Finish 1983-95 about Dec 1996

4. New radiation code
   — Will derive surface radiation from new clouds
   — Do radiation soon after clouds are ready

(Information is from Bill Rossow)

Roy Jenne
May 1995
Analysis Products
NMC and ECMWF


CD-ROM of NMC N. Hem daily grids, 1946-on

Regular NMC products

Advanced NMC

Reanalysis (NMC/NCAR)

Reanalysis by Dec 1995*

Regular ECMWF products

Advanced products

ECMWF reanalysis

* Maybe 40 years by fall 1997 (1957-96)

Roy Jenne
May 1995
Data for EOS Pilot Project
(Droughts 1988 and floods 1993)

- NMC analyses
  - and reanalyses
- ECMWF analyses
- Daily precip over N. America
  - from about 700 sites on real-time telecom (at NCAR)
  - from about 7000 sites over the U.S. (at NCAR, from NCDC)
- Hourly precip for 3000 stations for the U.S.
- Surface meteorology (wind, temperature, pressure, clouds, etc.)
  - Hourly data for 700 to 1000 stations
- River flow data
  - not at NCAR now
  - could get from USGS
- Some hourly surface solar data

Roy Jenne
May 1995
A New Dataset at NCAR
(SSMI products from Wentz)

- Received Apr 18
  - on 14 Exabyte tapes, guess about 65 Gbytes, cost $2800.
- Jul 1987 - Dec 1994 (14 satellite years)
- Two satellites from Jan 1990-on
- Variables
  - ocean surface winds
  - cloud water
  - total precipitable water
  - rain rate

Roy Jenne
May 1995