July 1999
Rev 17 Aug 1999

GLOBAL OBSERVATIONS FOR REANALYSIS, 1948 – ON
(Paper for 2nd International Reanalysis Conference)

Roy Jenne

National Center for Atmospheric Research, 1850 Table Mesa Dr., Boulder, Colorado, 80303, USA

jenne@ucar.edu

1. ABSTRACT

The status of preparing more observations for the next long reanalyses by ECMWF and NCEP will be reviewed. The coverage of global upper air, surface, and satellite observations is summarized for 1948 – on. The changing coverage of observations is related to the scores of forecasts made from reanalysis. These scores climbed to modern levels in 1973, in the Northern Hemisphere. A considerable number of digital observations are available starting in 1948. Observations first started in Antarctica and much of South America in 1957. The world's first satellite sounders started in April 1969, and were used in the NCEP/NCAR 50-year reanalysis from March 1975 – on. For the oceans of the Northern Hemisphere, there were more permanent ships taking rawinsonde observations during 1948 – 73 and more reconnaissance flights during early years, than in later years when satellite sounders were available. The additional old datasets and updates that became available after March 1998 are summarized. The main opportunities to add more data will be described. The number of global rawinsonde and pilot reports have been declining in recent years, which is a problem, while aircraft reports and satellite cloud winds have increased. NCAR has 500 mb daily grid analyses for the Northern Hemisphere for 1946 – on from operational maps. Maps of the differences between these grids and the NCEP/NCAR reanalysis will be presented for selected periods. Sources of more information about categories of observed data (rawinsondes, aircraft, etc.) will be presented along with sources of information about individual datasets. A limited history of the progress of operations during several long reanalysis projects will be given.

It has been possible to obtain and process many datasets which capture a large fraction of the world's observations that were taken during 1948 – 1998. These can be used for reanalysis projects and for other research. We will include text, tables and figures to give the reader a sense of the data coverage that is available.

2. THE WORLD'S RAWINSONDE OBSERVING NETWORK DURING 1940s – 1990s

The world's upper air rawinsonde network of stations was almost modern from July 1957 – on. Before that there were no observations for the West Coast of South America, none for Argentina, and data for Antarctica did not start until 1956 – 57. However, there are observations from Australia, New Zealand, parts of South America, and parts of Africa that often start by 1950, but the network may be sparse. There was a buildup of observations for the IGY "year" and those observations for about July 1957 – Dec 1958 appear to have been included in one or more of the input datasets used for reanalysis. Rawinsonde observations for the Northern Hemisphere and tropics generally were started by 1950, and some have earlier starting dates.

The political revolution in China was completed in 1949. A few raob stations for reanalysis are available starting 1954, but the good data coverage does not start until mid-1956. There are observations for India from 1950 – on.

The upper air observations for the early years have some advantages in the Northern Hemisphere compared with modern years. A number of permanent ships started rawinsonde observations in the late 1940s and data at these locations continued in the northern oceans until about 1973 – 74. After that, the number of ships was greatly reduced. There were weather reconnaissance flights over northern oceans and to the North Pole during about 1947 – 1959. These together with the permanent ship rawinsondes (~1948 – 73), and the island rawinsondes should help the analyses over the oceans in early years.

3. THE STARTING DATES FOR RAWINSONDE OBSERVATIONS

Many rawinsondes were taken in the 1940's especially in the No. Hemisphere. It appears that many of these early reports for the 1940s may be impossible to gather together. From about 1946 or 1948 and later, there are many observations in digital files that can be used for reanalysis. The start of rawinsonde observations in selected regions is shown in Table 1. It is now possible to make nearly all of this data available for reanalysis, after a large data preparation effort.

Table 1-Starting Dates for Rawinsonde Observations (Selected Regions)

Region	Starting Year	
Rawinsondes in Antarctica	1956, 1957	
Data for West Coast, South America	mid-1957	(But pibals start earlier)
Argentina	1957	(But pibals start earlier)
South Africa	1950	(But 1960 still missing)
Nairobi, Africa	1948	
Australia	1943 – 1949	
New Zealand	Mid-1949	
Remote islands, Southern Hemisphere	1950 – 1957	

4. COMMENTS ABOUT THE START OF RAWINSONDE OBSERVATIONS IN THE AMERICAS

a. US National Stations: The data on these tapes starts in 1948. Alaska, including the Aleutian chain of islands, has good data coverage starting 1948. Other tapes have data starting 1946, but these have not been converted as yet. The early data are in complicated card formats with overpunches. The US observing net of aircraft temperature soundings or radiosondes is from 1928 – on, with more stations in 1935. But the data prior to 1946 are no longer in digital form.
b. Canada: Seven or eight stations in far north Canada are available from 1948 – on. Data for Southeast Canada is also available from 1948. Hudson Bay starts 1950. Raobs for the rest of Canada start July 1955. There were earlier observations but they have not been digitized.
c. Mexico: Three stations started in 1948, about two in the 1950's and more later.
d. Caribbean: The first four stations started in 1947, 1948, 1950, 1953. Six more stations started in 1956 or 1957.
e. Panama Canal Zone: The rawinsonde data are for November 1946 – on.
f. West Coast of South America plus Argentina: Raob observations started July 1957. And there was earlier pibal data.
g. Data for Brazil: There is little data for the 1940's. NCAR obtained data directly from Brazil. There were 11 raob stations (the earliest started 1961, 63, 65, 66, etc.) There were 29 wind-only stations; two started 1950, 8 started 1951, etc. This data ended about 1986. We will try to obtain an update.

5. RAWINSONDE DATA COVERAGE FOR SELECTED ASIAN AREAS.

a. China: We have data for a few raob stations for 1954 – 56. The period of very good coverage for China starts September 1956 (about 100 stations in TD54).
b. India: NCAR received data directly from India for the period 1950's – 1978.
c. Singapore: The continuous record at NCAR starts 1957. The data came from two or more sources.
d. Hong Kong: We have raobs for November 1949 – on. April 1954 – March 63 came from UK; 1956 – 90 was from NCDC; Nov 1949 – Mar 54 was from raob seq. 19 at NCAR. TD54 has part of the data from 1950 – on.

The coverage of rawinsonde data for reanalysis for two sample years, 1950 and 1958, is shown in Figures 1 and 2. There is some additional data not shown in these plots.

6. PIBAL DATA COVERAGE FOR ABOUT 1942 – 1966

Winds aloft from the radar or visual tracking of balloons (called pilots or pibals) are an important source of data for reanalysis. For early years, there are millions of these observations, especially in datasets TD52 and TD53. The TD52 dataset has 3,786,205 wind aloft observations and TD53 has 6,107,924 reports. NCAR has texts and coverage plots that show the annual coverage of these data from TD52 and TD53 for each early year. There is considerable data coverage for places like Africa, South America, India, US, etc. In this text we will focus on the years 1942 – 1966. Figure 3 shows the coverage of pibal wind observations during 1945, 1949, and 1954.

Data archives of upper air data (raobs, pibals, aircraft) from the NCEP decode of GTS data start in March 1962. The coverage of that data is not show here. For reanalysis, we also had pibal data for early years received directly from several countries, including Australia, Brazil, and France. That data (mainly for about 1950 – on) is not on these plots. Figure 4 shows the average daily count of pibals for each year during 1942 – 1990. There is a dip in data counts during 1964 – 69. The old data preparation methods were terminated before the new methods were adequate.

We also remind the reader that data from a rawinsonde ascent is a more powerful observation than a pibal. A raob has wind, temperature, and pressure information, and usually goes higher than a pibal. The data coverage of rawinsondes is discussed elsewhere. But pibals (or pilots) give winds, a very important variable to help reanalysis.

The data coverage plots of pibals are encouraging (Figure 3). The rawinsonde observations for Argentina and the West Coast did not start until Jul 1957, but the pibals started 1946 or 1948. I am pleased that at least this much of the older data did get put into digital form, and did survive until now. Data retention and data survival are worse problems than most people realize.

7. COMMENTS ON THE COVERAGE MAPS OF PIBALS FOR 1942 – 1966.

- The USA-48, Alaska, Hawaii, Caribbean, and Central America generally have good coverage for 1942 – 1964 and similar data will be on the NCEP tapes with GTS data starting in 1962 (not shown here).
- Data coverage for Argentina is good for 1946 – 58 and 1963 – 1967. Venezuela is good during about 1951 – 1967.
- The Latin American West Coast is good during most of 1948 – 67, but weak during Jan 54 – Jun 57.
- Ten stations in Brazil started in 1950 or 1951 in another dataset (from Brazil), not plotted here.
- Greenland has data coverage during 1942 – 1959.
- Northern Canada has good observations here for 1948 – 1958. Then similar observations should be in other datasets.
- All of Canada has good pibal data coverage here for mid-1955 through 1958.
- India has good coverage for 1944 – 48 and for 1953 – 67. We are also using a dataset of rawinsonde data from India for 1950 – on.
- Southeast Asia has good coverage during 1953 – 1967.
- Japan has very good coverage on these charts during 1942 – 1957.
- Data coverage of pibals for Japan (more information):
  - Coverage for 1925 – 1953 (in TD52)
  - Excellent coverage for 1931 – 1953 (in TD52)
  - Good coverage (or excellent) for 1945 – 1957 (in TD53)
- China has reasonable coverage for 1943 – 47 and excellent coverage for 1956 – 59. We also note (from other information) that the coverage of rawinsondes starts in 1954 in China and is very good from mid-1956 – on.
- The Mideast has fair coverage during 1949 – 1959.
- Africa: The data coverage for Africa is helpful for 1949 – 1966. But there is no data for the southern part of Africa during the whole period (but there is some raob data). Maps are available to show how the data coverage changes over different parts of Africa. (Figure 3, etc.)

8. AIRCRAFT DATA FOR REANALYSIS

A considerable amount of aircraft and recon data is available for reanalysis. The USAF flew several long recon tracks over the Atlantic and Pacific Oceans during 1947 – 59. The flight level data was digitized (dataset TD57). So far, the dropsonde data has not been prepared. Some of the recon tracks were flown once per day. The Arctic Ptarmigan track from Alaska to the Pole was less frequent.

Jim Sadler, University of Hawaii, worked on projects to collect tropical aircraft data for a number of years. The data is for 1960 – 73. I think that some data from NCEP/NCAR was included. He prepared the data and drew maps. All of his observations are being used for reanalysis (Table 2). The GATE tropical Atlantic experiment was held over the ocean west of Dakar, Africa, during Jun – Sep 1974. All of the data from ships and aircraft is being used in reanalysis.

Aircraft data reported over GTS telecom in real time is being used for reanalysis. The data collected by NCEP and archived at NCAR starts in March 1962, see Figure 5 (now 37 years of observations). The Navy source has higher data counts than NCEP during early years, so that data will also be made available.

Table 2-Aircraft Data for Reanalysis

Dataset	Data Years	Remarks
TD57	01/47 – 01/59	USAF Recon, Northern Oceans and Arctic
Sadler	1960 – 1973	Tropical Aircraft
Gate	06/74 – 09/74	Gate Tropical Atlantic Experiment
Gasp	1975 – 1979	NASA Tropical, Wide-Body Aircraft
FGGE Year	12/78 – 11/79	Used for Reanalysis
NCEP, GTS	03/62 – 12/98	Now Data for 37 Years. Used.
USAF, GTS	1976 – 1985	Have used 1976 – 78
Japan, GTS	Mid/78 – 12/94	Used for Reanalysis
Navy, GTS	1971 – 1990s	Not Used Yet
Australia, GTS	12/71 – 11/89	Ready to Use in July 1999.
NZ Acft, GTS	12/78 – 06/88	Used for Reanalysis

Australian Aircraft Data (1971 – 89): The aircraft data from Australia has many reports that never reached the NCEP data files. The Australian data will be ready to use in July 1999. The counts of reports and the rate of duplication with NCEP are given in Table 3.

Table 3-Australian Aircraft Counts, and Duplication with NCEP
Counts of Australian aircraft for 1971 – 89 are given for selected years.

	Total Count	Counts per Day	Dupe Rate with NCEP	Approx Counts Non-Dupes
1975	25,157	69	9%	22,870
1978	25,740	71	51%	12,480
1984	95,668	262	90%	9,180
1985	255,298	699	59%	105,900
1989	1,131,351	3,100	68%	358,600

9. CONSTANT LEVEL BALLOONS FOR THE SOUTHERN HEMISPHERE

There were two large experiments in the 1970s, during which many balloons were released that drifted in the winds at levels near 150 millibars over the Southern Hemisphere. During some periods, about 100 balloons were drifting at a given time; therefore they described the circulation of the Southern Hemisphere at 150 mb quite well. The TWERLE balloons even had a small accurate altimeter, so that the observations at altitude were height, pressure, temperature, and wind (derived from location). TWERLE data was used for the NCEP/NCAR 50-year reanalysis and will be used for ERA-40. NCAR has archives of EOLE balloon locations, but not of derived winds. The periods of data are:

	Period of Data	Comment
EOLE	08/1971 – 07/1972	Not used yet
TWERLE	07/1975 – 08/1976	Used in NCEP/NCAR

10. DATASETS FROM SATELLITE SOUNDERS (APR 1969 – 1999)

The first satellite sounder data started in Apr 1969. There may be a few months of gaps in the early few years of sounder data, but by the end of 1998, the world had almost 30 years of these data. Satellite soundings help to give temperature data for the atmosphere over the broad ocean areas where there are no rawinsonde data except for a few islands. Table 4 shows the availability of satellite sounder data. In the present NCEP/NCAR 50-year reanalysis, sounder data was used from March 1975 – on. The quality of the earlier derived soundings was too poor to use, and there was not time to develop procedures to directly use radiances for either VTPR or for SIRS (see Table 4). In ERA-40, the VTPR radiances will be used.

The selective chopper (SCR) data from Nimbus 4 and 5 was mainly designed to sample the stratosphere, but it does include tropospheric data. Therefore, we hope that it will help to bridge the gap between SIRS data and the VTPR data that starts Nov 1972.

Table 4-The Main Datasets from Satellite Sounders (1969 – 1999)
Level 1 (L1) means basic radiances. Level 2 (L2) means soundings.

		Data	Years	Remarks
a.	SIRS A&B, L1	69/04 – 71/01	1.8	Radiances
	SIRS (NCEP), L2	69/11 – 72/09		Soundings
b.	IRIS-D L1	70/04 – 70/12	0.7	Looks down, not scan
c.	SCR (Nimbus-4)	70/07 – 73/01	2.6	
	SCR (Nimbus-5)	72/12 – 74/12	2.0	4.5 years of SCR
d.	VTPR (NOAA), L1-2	72/11 – 79/02	6.3	8 IR channels
	VTPR , NCEP, L2	73/01 – 79/02		Soundings
e.	TOVS 2.5^{o}, L1.5-2	79/01 – 94/01		
	TOVS, L1 (HIRS,MSU, SSU)	78/11 – 98/12	20.2	IR, VIS, microwave, strato.
f	UARS	91/09 – on		On CD at NASA

Note: The derived VTPR soundings were not good prior to Mar 1975. The VTPR radiances should be good for the whole period.

11. CLOUD WIND DATA FROM SATELLITES (1963 – ON, MOSTLY 1973 – ON)

Cloud winds: There are a few blow-off satellite winds starting in 1963. These were estimated by looking at the direction of the movement of clouds blowing off of the tall cumulus clouds. The more precise cloud drift winds start in 1967 for the region of the Americas. These are based on deriving cloud movement during the half-hour period between successive pictures from geosynchronous satellites. By the late 1970's, there were similar satellites operated by Europe and by Japan. These data are important sources of wind information over the oceans. The wind data is both for low cloud levels and high clouds.

The first geosynchronous satellite (ATS-1) was launched by the US in Dec 1966. A second satellite (ATS-3) was operational from Nov 1967 – Dec 1974. ATS-1 was located about 151^{o} West and ATS-3 was between 45^{o} W and 95^{o} W. But we do not have very many satwind observations to use during 1967 – 72.

The number of satwind reports available to use for reanalysis is about 82 reports per day in 1967, and 120 to 190 per day during 1968 – 72. For 1973 – 75, the counts increased markedly to 372, 482 and 1105 reports per day, respectively (Figure 5). The 5-day forecast scores in the Northern Hemisphere from reanalysis also increased to high levels in 1973 – 74 (Figure 11) and this added cloud wind data was probably one of the factors that helped.

The distribution of cloud wind data by year and by longitude is shown in Figure 6. The chart is based on cloud wind data used in operations at NCEP. It shows average data counts for each 20-degrees of longitude and for each 3-month period. During the FGGE year (Dec 78 – Nov 79), there

was also a satellite over the Indian Ocean and data was processed in delayed time. It got into FGGE datasets and all of that data is being used in the reanalysis projects. There is a data gap during 1980 – 82 in the Europe – Africa sector (Figure 6). We hope that reprocessing efforts by Europe may fill that gap.

A project to prepare new Meteosat winds is in operation (written March 1999)

First, Europe will reprocess Meteosat 2 data (1981 – 1988). This will be done between July 1999 and July 2000. In mid-2000, they may start a project to reprocess Meteosat-1 data (for 1977 – 1979). The ERA-40 reanalysis of 1979 – 1989 will not start until the new Meteosat winds are available. They now (Mar 1999) are copying Meteosat-2 data for 1981 – 1989 onto DLT tapes (volume 750 Gbytes/year). This is nearing completion. But there were media problems on 10% of the tapes that they have to take care of.

The FGGE year: As noted above, there was good coverage of cloud winds for the FGGE year (Dec 1978 – Nov 1979), and these data are being used by the reanalysis projects. During FGGE, the US had two geosynchronous satellites viewing the Americas, Europe had one to view Europe and Africa, and Japan had one. Also, the US operated a satellite over the Indian Ocean during FGGE. A separate text about cloud winds and data coverage is available.

12. OBSERVATIONS AT THE EARTH'S SURFACE

Over the oceans we are using the COADS dataset which is the world's best collection of surface marine data (from ships, fixed buoys, drifting buoys, arctic ice buoys, etc.) For the land areas, we have datasets of data from GTS for 1967 – on. Other datasets give us a considerable amount of data coverage for world land areas from the 1940's and later.

13. SURFACE OBSERVATIONS

Surface synop observations based on GTS telecom data are available from 1967 – on (now 32 years). The data for surface observations came mainly from the NCEP decode of GTS data for July 1976 – on, and from a USAF dataset for 1967 – 1976 based on GTS input. Reports are typically obtained from about 7500 land stations in the world. The data for the earlier years came from the TD13 dataset, which has significant coverage for the world, but not complete coverage. We also used data for 300 US stations (from NCDC) for early years (1948 – 67) as well as 223 stations from Russia (for former USSR), 1948 – on. It would be desirable to add more surface data from other regions of the world for early years. However, the available coverage plots at NCAR show that a considerable amount of data coverage already has been used by the reanalysis projects. For the global oceans, the COADS dataset was used for ship and buoy data. COADS is a cooperative project between NCAR, NOAA-ERL, and NOAA-NCDC. Other organizations such as MEDS, Canada, the UK Met. Office, and Russia have also helped. Many countries have contributed their marine ship and buoy data.

14. SURFACE DATA COVERAGE OVER RUSSIA AND ASIA

We will summarize the coverage of surface data in the TD13 dataset from the USAF.

Table 5-Data Coverage of Surface Synoptic Data for Russia and Asia
Data coverage from the TD13 dataset prepared by the USAF.
China and Russia have good data coverage in other datasets starting Jan 1967.

Year	Mid-Russia	Hi Lat Russia	China	India
1946	None	None	None	Dense
1948	Good	Poor	None	Dense
1954	Dense	Poor	None	Dense
1956	Dense	Good	Dense	Dense
1963	Dense	Good	Dense	Dense
1967	Good	Good	None	None

NCAR has a dataset of USSR surface synoptic data from Russia for 1936 – on. We used this for 1946 – 1956 when the coverage of stations in northern Russia is not very good.

There are other tapes with 3-hourly surface synoptic data (about 7500 stations for the world) that give good coverage starting Jan 1967. Thus, we should not worry when there is poor data coverage for 1967 in the table above.

15. COVERAGE PLOTS OF SURFACE SYNOP DATA

The coverage of surface synoptic data for reanalysis over Latin America is given for 1977 in Figure 7. This is data from the NCEP decode of GTS and the reanalysis data coverage for all years 1967 – on should be similar to this. A similar data coverage plot for Africa is shown in Figure 8. Data coverage for Africa for 1948 – 1967 from dataset TD13 is in Figure 9. Also, world coverage for TD13 for 1948, 1964, and 1970, is in Figure 10.

16. SURFACE DATA COVERAGE OVER THE OCEAN, 40 – 60^{o} SOUTH

For the one-year FGGE experiment (started December 1978), drifting buoys with pressure sensors were deployed in the west wind belt of the Southern Hemisphere. This array of buoys died, but it revived about 1984 and still continues (1999). In early years, there was data help from whaling ships, especially in the southern summer. Harry van Loon thinks that the daily Southern Hemisphere sea level pressure maps (drawn in South Africa) are rather good starting in December 1953. These were published in NOTOS through April 1964. See Taljaard and van Loon, 1964, for a description of the IGY analyses *(Bull. AMS,* Feb 1964). Some of the gridded data are at NCAR.

17. WHAT IS NEW FOR OBSERVATIONS AFTER MARCH 1998

NCEP obtained (from NCAR) the last of the observations needed to do reanalyses for 1948 – 57 in early March 1998. NCEP started production on these years on 20 Mar 1998 and finished 23 Jul 1998. TD90 is a dataset of about 30 raob stations for 1943–62. NCEP got these data, but they did not have time to sort them into the production stream. They will be used for ERA-40. The C-Cards dataset of raobs for 1949–66 is from the old NCDC Northern Hemisphere data tabulations project. NCEP used C-Cards data for 1949 – 59; ERA-40 will have a better version of the whole period (1949 – 66). The version for NCEP was quite good, but NCAR found many more cases where the units for some levels and some stations had irregular dates of change. These have been fixed. There are French raobs for 1948 – 79. NCEP used 48 – 57; ERA-40 will have all, plus additional wind-only data for 1964 – 79.

The NCEP upper air data for 1962 – 72 from GTS had cases where the same observations were repeated for different date/times. This was probably due to a computer disk buffer that did not get any new data, but it was given a new date/time. NCAR removed 18 of these cases before NCEP got the data. Some additional, more complicated, versions of this problem were found and removed later. TD54 is a large dataset of global raobs for 1943 – 68 that has fewer observations in early years. NCEP got 1943 – 67, not 1968. ERA-40 will also have 1968. Station 55299 for 1 Jun 57 – 27 Dec 59 had surface pressure in the surface height slot. This is fixed for ECMWF.

Aircraft data from Australia for 1971 – 89 was not used by NCEP, but is available for ECMWF. The duplication rate with NCEP data is only 10 to 50% in early years; therefore this data will add to coverage. The University of Wisconsin runs an automated network of surface stations for Antarctica (1980 – 99) and Greenland. For NCEP the data went through 1992. Recently the data has been updated through at least 1996. NCAR ran many more checks. Wisconsin has corrected parts of the new and old data. It is a nice dataset—with stations in very remote regions. Russian 3-hourly surface observations are available from North Pole ice islands for about 38 years: 04/1951 – 03/1991, and almost complete for 04/1954 – 03/1991. Certain problems in the location and date/time of observations are almost solved. The data will be available in Aug 1999.

NCAR has done a lot more work on upper air station libraries (the location and elevation of stations). The early tapes of operational GTS data had a number of stations that were at wrong locations. NCAR reduced the errors enough before NCEP got the data, that the NCEP/NCAR 50-year

reanalysis should not have been significantly hurt by poor information. The ERA-40 will have the benefit of two or three more rounds of error reductions in library data.

18. DIAGNOSTICS TO DETECT ERRORS IN DATASETS OF OBSERVATIONS

NCAR concentrates on developing programs to detect and fix (when possible) the errors that will cause large problems for reanalysis, and which can not be handled well by the quality control programs in the centers. We also concentrate on trying to correct the station libraries, so that the data will be located in the correct place. A few hundred rawinsonde stations were misidentified in one dataset. This puts good observations at a wrong location. In one early dataset (US Control raobs) there were 76 stations with 121 station-years of rawinsonde data that could not be used. The date/time of the data was wrong (usually by 12-hours), and data from different times was mixed together in each sounding. Almost all of this data could be recovered from another dataset (TD54). These stations included San Diego and Kwajalein Island. Two stations had daily data for several years, but it was all the same observation. Another dataset (TD54) had the stratospheric raob data wrong for 61 stations. The data were assigned to the wrong pressure levels. Most of this problem could be corrected by using the US Control raobs. We have made massive comparisons of actual observations between datasets that have some overlapping data. This is one of the methods that have helped us to identify the problems. In most cases, but not all, the problem can be fixed. At least we know how to avoid the bad data. There are also thousands of reports that have various types of random errors. We have fixed some of these, but these are mainly handled at NCEP or ECMWF, which have very capable quality control programs.

In a few cases, we found that rawinsonde data had been assigned to wrong levels in parts of some datasets. Sometimes, only parts of each sounding were systematically at the wrong levels. One method that we used to detect these problems was to do hydrostatic checks on all of the rawinsonde data, and then accumulate monthly statistics about the failure rate for each level, and each station.

19. DEVELOP AN UPPER AIR STATION LIBRARY

It is important to know the location and elevation of the world's observing stations so that the data can be used in the proper way. There are some existing source libraries (such as CARDS, WMO listings, and USAF tapes) but no one source has information about all stations over the past 50 years. Also, there are some mistakes in the libraries. Some stations have moved and it is very difficult (sometimes impossible) to find information about all of these changes. NCAR digitized some information from old source documents, as for the TD52 pibal dataset, and the C-Cards rawinsonde dataset. NCAR has compared different library datasets, and prepared time histories of the data on operational tapes to help remove errors.

Version 1 of the upper air library was used for data sent to NCEP during 1993 – 03/98. It was developed during the 1993 – 97 period. Version 2 with lower error tolerances was ready by Oct 1998, to insert into the upper air reports sent to ECMWF. Version 3 will be ready by Sep 1999.

20. SEVERAL REANALYSIS PROJECTS

Brief information about five large reanalysis efforts is listed below. The reanalysis projects by NASA, Goddard are not included in this list.

a. Data for NCEP/NCAR 50-year reanalysis (1948 – 97)
   - Production started 06/1994 and finished 07/1998.
   - Last date for model changes was 08/1994.
   - T62, 28 levels; 208 km resolution.
   - The dimension of the Gaussian Global grid for T62 is 192 x 94 points.

b. ERA-15 by ECMWF (1979 – 93)
   - Production starts 06/1994 and finished 09/1996.
   - Last date for model changes was about 11/1994.
   - Model resolution T106, 31 levels. Resolution 1.125 degrees or 125 km.

c. A new NCEP reanalysis of 1979 – 1997

- Production starts May 7, 1998 (T62, 28 levels, 208 km).
- Major model changes completed May 1998.
- Aug 21, 1998: Now doing Mar 1983 (1979 – 82 done).
- Feb 22, 1999: Now doing Dec 1988, going smoothly.
- Jul 1, 1999: Now doing Jun 1992

d. ERA-40 by ECMWF (1957 – 2001?)
   - Production may start by Dec 1999.
   - Production will take about two years, using two operational data streams.
   - Probably T159, 60 levels; resolution 0.75 degree or 83 km.

e. New long reanalysis by NCEP (probably for 1948 – 2003)
   - Production may start in year 2002.

21. AVAILABLE OUTPUT FROM REANALYSIS

The NCEP/NCAR 50-year reanalysis has many outputs as described in the Mar 1996 *Bull. AMS* paper (Kalnay, et al, 1996). The most used outputs are 1) annual CD-ROMs, 2) monthly statistics (0.6 GB/y), 3) data on 17 pressure levels (1000 to 10 mb), at 2.5^{o} resolution, and 4) flux data (precip, radiation, clouds, etc.) The volume of the 2.5^{o} pressure data is 2.53 GB/y and flux data is 1.99 GB/y. The total archive has about 54 GB/year.

The new NCEP reanalysis (1979 – 97) uses a model updated to May 1998; the data will be made available by NCAR and NCEP in a similar manner.

The output from ECMWF (ERA-15 and ERA-40) is similar in data types to the NCEP output. It will be available from ECMWF and US research users will also have access via NCAR.

22. FORECAST SCORES RELATED TO AVAILABLE OBSERVATIONS

During the long reanalysis for 1948 – 98, NCEP ran one 8-day forecast each fifth day. NCEP has prepared plots of the annual 5-day forecast scores for each hemisphere, shown in Figure 11. It is difficult to make an accurate 5-day forecast, but even these 5-day scores are quite good, especially in the Northern Hemisphere. We are pleased that the forecast scores from reanalysis are at modern levels for the Northern Hemisphere during 1973 – 98. The scores during this modern period (about 0.71 correlation) are much better than the NCEP operational scores used by the public in 1984 (about 0.55). The reanalysis scores for all of 1952 – 1998 are better than the operational scores in 1984. These high forecast scores for reanalysis give us confidence that the quality of the analyses should be high.

We think that the drop in forecast scores during 1971 – 72 in the Northern Hemisphere was caused by a drop in observations. We believe that the operational data ingest at NCEP was not maintained well during these last years of the old data ingest system. We plan to see if Navy (Monterrey) observations will help with the supply of data during 1971 – 72. NCEP started using a new data ingest system in Jan 1973. It lost the 06 and 18 GMT rawinsondes for two years, but this did not seem to hurt forecast scores much (Figure 11), if any. There may be a way to recover this data from another dataset from the USAF.

The MIT dataset of raobs is for May 1958 – Apr 1963 (5 years). The forecast scores did not drop either before or after this period. Therefore most of the MIT observations were probably also carried in other component datasets. The Northern Hemisphere forecast scores dipped down in 1955. This may be explained by several months of missing data in the C-Cards dataset in 1955.

Satellite wind reports are based on the movement of clouds in successive satellite pictures. The number of cloud wind observations for reanalysis were about 120 to 190 per day during 1968 – 72. Then the counts increased markedly to 372/day in 1973, 482 in 1974 and 1105 in 1975. This increase in wind data coincides with better forecast scores for 1973 – 75 than for earlier years. I do not know why there were so few wind observations during 1967 – 72. The first geosynchronous satellite, ATS-1, was launched in Dec 1966, and ATS-3 was operational during Nov 1967 – Dec 1974.

The world's satellite sounders started in April 1969 but they were first used in the NCEP/NCAR reanalysis in March 1975. The Northern Hemisphere forecast scores are almost equally good for 1973 – 98. Therefore, it is likely that VTPR soundings only helped the Northern Hemisphere over portions of the Pacific Ocean. Parallel reanalysis tests were made for Jan – Feb 1979 using VTPR, TOVS, and nosat cases. The differences between TOVS and VTPR analyses were small, but nosat was quite different over remote ocean areas. Sounders help the placement of daily circulation patterns, especially in the west wind belt of the Southern Hemisphere.

The forecast scores for the Southern Hemisphere became relatively good during about 1985 – on, but significantly worse than the Northern Hemisphere. When TOVS radiances are directly used (not the soundings) for the next reanalyses, we hope that the Southern Hemisphere scores will look better for the whole 1979 – 99 TOVS period and perhaps for the whole period starting fall 1972 when the VTPR sounders started.

Australia prepared bogus point data of sea level pressure taken from their maps of analyzed pressure. These were used at the wrong location (180 degrees of longitude off) during Jan 1980 – Mar 1993 in the long reanalysis. The first guess checks threw out half of the data, but some bad data was used, and the analysis did not have the benefit of using the data in the right location. This probably hurt the Southern Hemisphere forecast scores during this period. However, the year 1979 had the benefit of all the FGGE data and the SLP bogus was properly used, yet it does not have high forecast scores.

23. BACKGROUND INFORMATION ABOUT DATASETS

Information about many datasets of observations and analyses is given in "Data Sets for Meteorological Research (Jenne, 1975). Another text (62 pages) was prepared to summarize sources of data that could be used for reanalysis (Jenne, 1991). A number of data inventory plots are in Jenne, 1992, that give a feeling for the coverage of different types of observations for reanalysis. When the possibility of conducting reanalysis projects was being discussed in the late 1980's, I wondered whether the checking of the output analyses could be automated enough that many days of analyses could be prepared during each day of operations. GFDL had good success in automating the reanalysis of the FGGE year (Dec 1978 – Nov 1979). This was good evidence that the amount of human time needed to carry out a long reanalysis could be kept under control. I prepared a paper that summarized available datasets for reanalysis and also had this operational information (Jenne, 1988).

24. SELECTED ADDITIONAL TEXTS ABOUT DATA FOR REANALYSIS

Many texts have been prepared to help document the datasets that are being used for reanalysis. Some of the texts have information about the progress and history of the NCEP/NCAR project, and they contain information about the output datasets. There is a list of papers about these subjects (over 80 by 1999). Some texts need revisions because we are learning new things rapidly. A few of the titles follow:

Text	Date	Pages
Initiative to Prepare Data Inputs for Reanalysis	4 Mar 1991	62 pp
Data for Reanalysis, Inventories	Nov 1992	115 pp
Rawinsonde Data for Reanalysis	24 Oct 1994	
Rawinsonde and Wind Data from Countries	21 Jan 1997	17 pp
The TWERLE Balloon Dataset	8 Apr 1996	
Global Satellite Sounder Data	Aug 1994	54 pp
Ice Cap Buoy Update	5 Apr 1994	15 pp
Possible Manual Map Help for Reanalysis, 1957 – 72	25 Mar 1996	7 pp
Status of Reanalysis	Oct 1995	38 pp
Selected Counts of Reports for Reanalysis	10 Jun 1997	9 pp
Several Reports, Data Counts and Data Coverage	Mar 1999	45 pp

Our plan at NCAR is to prepare more documentation about the datasets used for reanalysis, about data coverage, and about the reanalysis project. Then we will place many of the documents and original manuals on a CD-ROM.

25. HOW TO ACCESS DATA COVERAGE AND INVENTORY INFORMATION

Several texts are available that give data coverage and data count information in different datasets. Also selected inventory counts are maintained on-line. See www.scd.ucar.edu/dss/ for a list. Examples are:

- Annual counts of NCEP Upper Air Observations by type, 1973 – 98.
- Monthly counts of raobs plus pibals for each station in NCEP data 1973 – 98.
- Annual coverage maps of pibals in TD52 and TD53, 1940 – on.

26. ACKNOWLEDGEMENTS

The project at NCAR to prepare 50 years of observations for global reanalysis has been operational during 1991 – 99. For several years it was a very intense effort to stay ahead of the operational needs of NCEP. In our Data Support unit at NCAR, several people have worked very hard to prepare and check the datasets: Dennis Joseph, Steve Worley, Chi-Fan Shih, Wilbur Spangler, Bob Dattore, Joey Comeaux, Gregg Walters, and Roy Barnes. Also, we say thanks for our administrative help from Cecilia Banner.

Several large datasets for early years were prepared by either the USAF or NCDC sections at Asheville and have been very valuable. Richard Davis was the liaison to this project from NCDC. Ernest Kung (University of Missouri) provided the MIT rawinsonde dataset and did recent work on it. John Lanzante from GFDL did the initial stage of processing on 70% of the large TD54 rawinsonde dataset. A number of countries directly provided data that help reanalysis. Several detailed lists are available.

People in many countries and laboratories also deserve credit for data preparation. Key participants are the observers around the world who work day and night to take the observations that make these analyses possible. Their contributions are gratefully acknowledged.

It has been a pleasure to work with the reanalysis teams at NCEP and ECMWF on these difficult projects. Eugenia Kalnay was the NCEP leader of the NCEP/NCAR 50-year project. Bob Kistler did data work and handled production. Jack Woollen handled the processing and conversion of data inputs. Our main day to day contacts with ECMWF to handle similar problems have been Adrian Simmons, Sakari Uppala, and Rex Gibson.

The project has been supported since its inception by the NOAA Office for Global Programs, and by the National Science Foundation. Without their enthusiastic support, we would not have been able to develop this project.

References:

Jenne, Roy L., 1975: Data Sets for Meteorological Research. NCAR TN/IA-111, Boulder, CO, 194 pp.

Jenne, Roy L., 1988: A Global Reanalysis; Data Inputs and Methods. Proceedings of the 13th Annual Climate Diagnostics Workshop, Oct 31–Nov 4, 1988. NOAA Climate Analysis Center/NMC; pp 79–84.

Jenne, Roy L., 1991: Initiative to Prepare Data Inputs for Reanalysis. An NCAR publication. 62 pp.

Jenne, Roy L., 1992: Data for Reanalysis; Inventories. An NCAR publication. 115 pp.

Kalnay, et al, 1996: The NCEP/NCAR 40-Year Reanalysis Project *Bull. AMS,* Vol 77 No 3, p 437–471

Taljaard, J.J., and van Loon, H., 1964: Southern Hemisphere weather maps for the International Geophysical Year. *Bull. AMS,* Vol 45 No.2, pp 88 – 95.

Rawinsonde Coverage

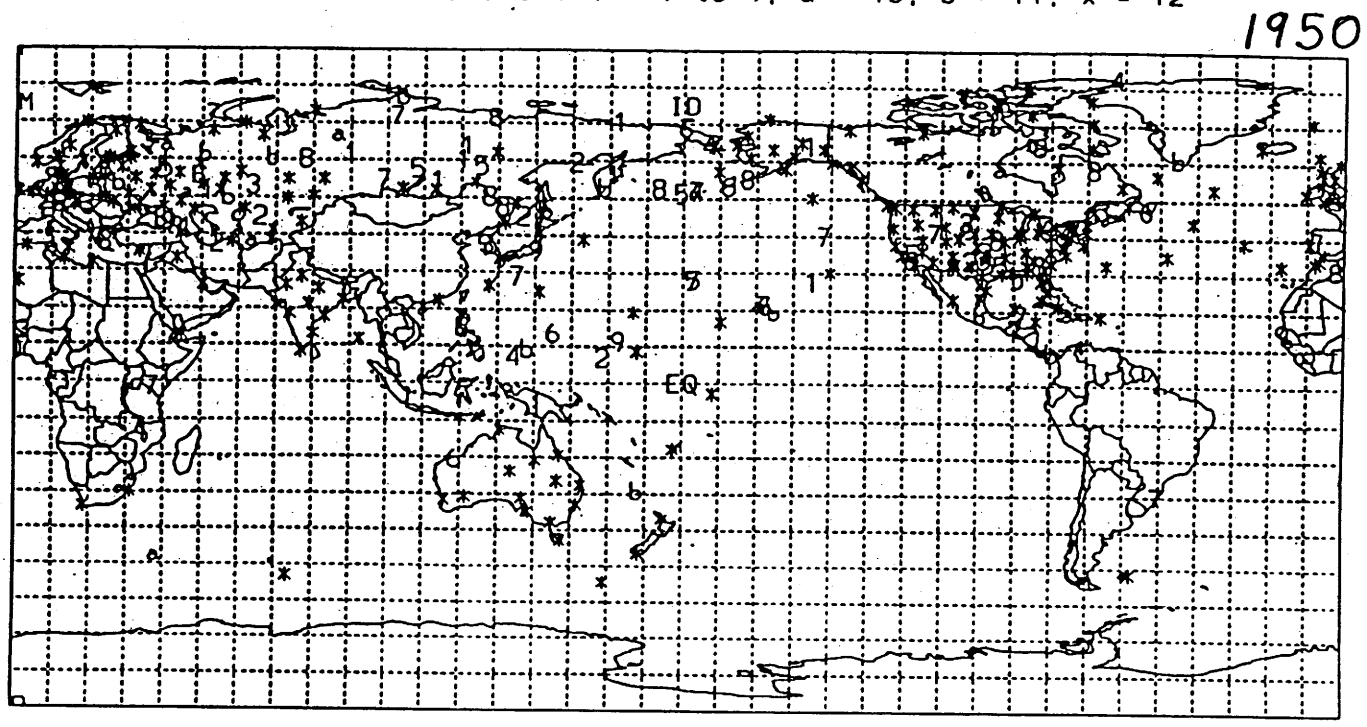


Figure 1. Daily rawinsonde coverage for 1950. A * means daily data for the full year; 7 means seven months. This shows much of the data for reanalysis, but not all.

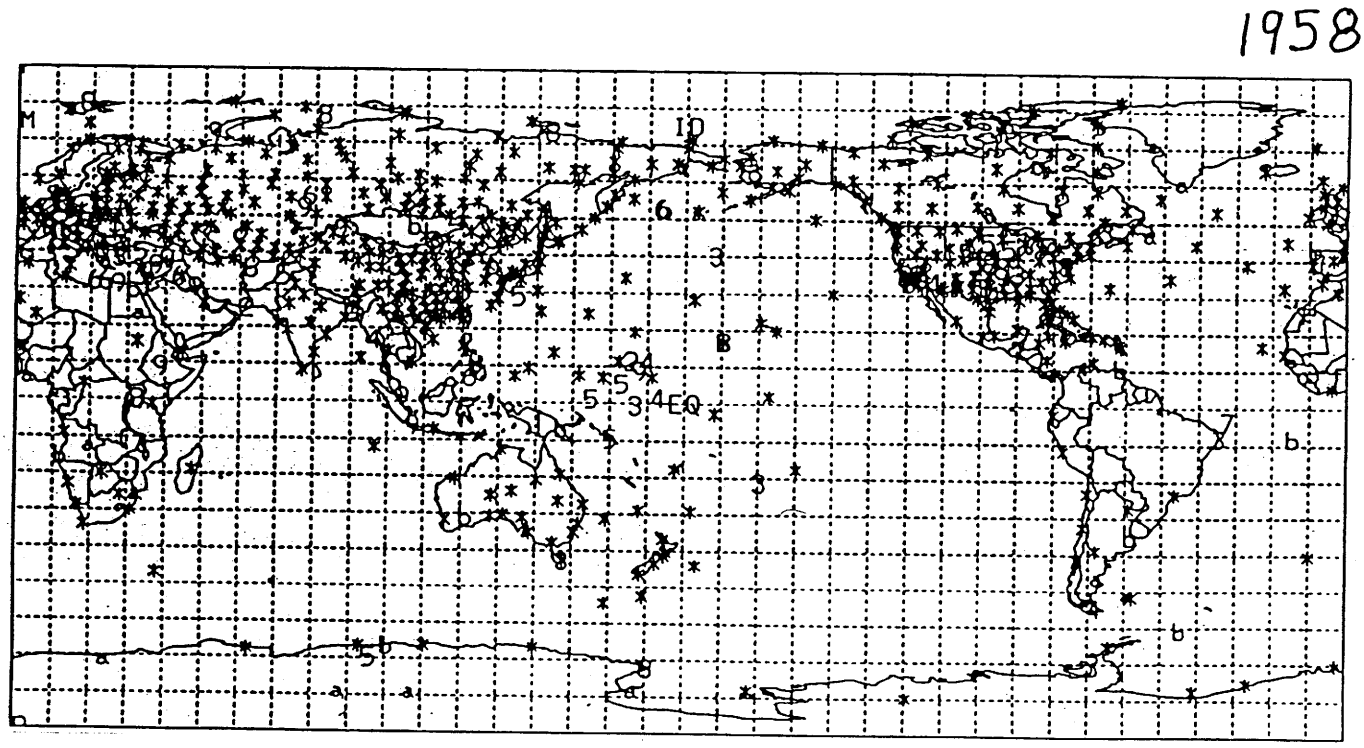


Figure 2. Daily rawinsonde coverage for 1958. A * means daily data for the full year; 7 means seven months. This shows much of the data for reanalysis, but not all.

Pibal Wind Coverage

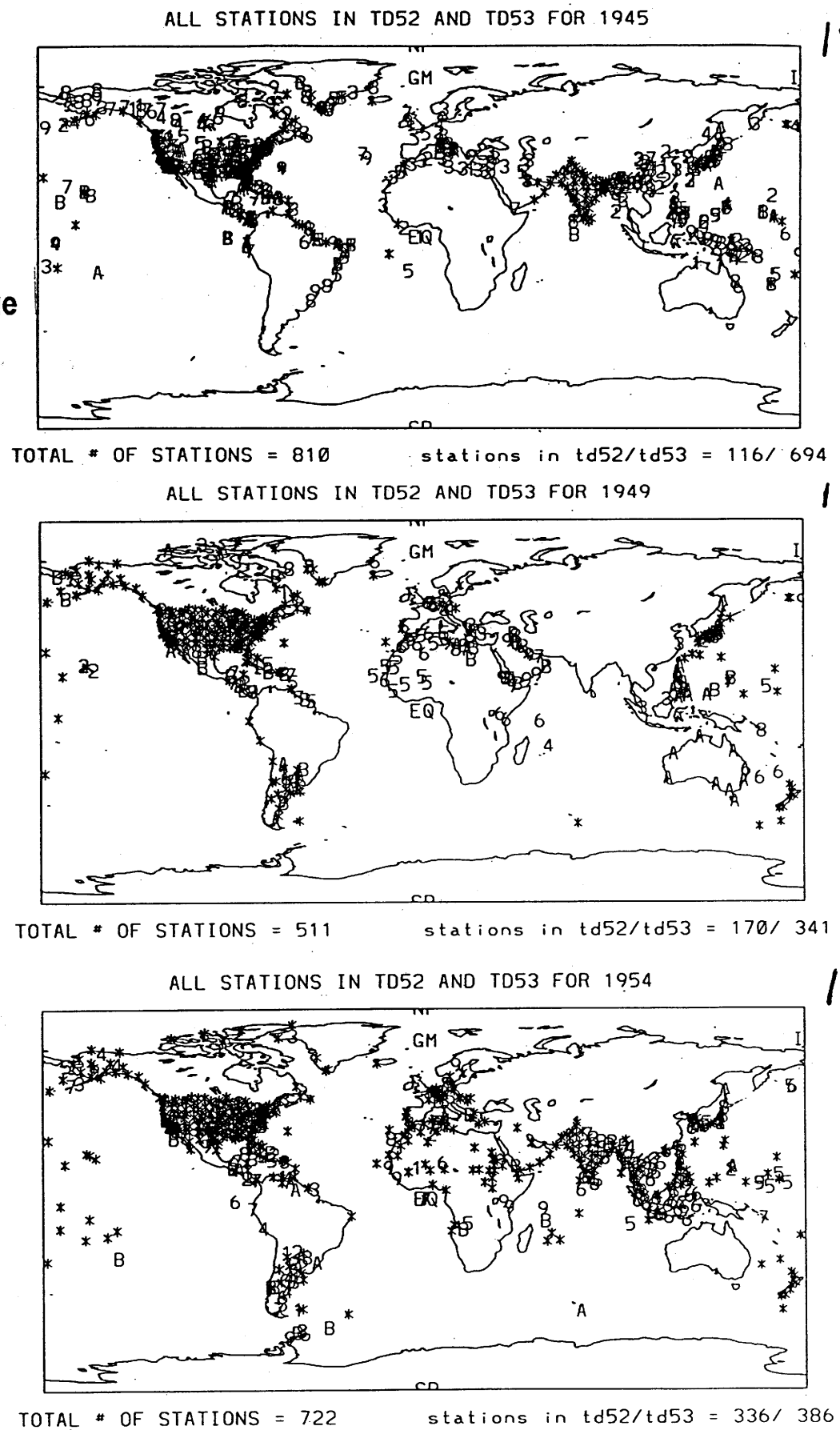


Figure 3. Daily pibal wind coverage for 1945, 1949, and 1954 from datasets TD52 and TD53. Also, 10 stations in Brazil started in 1950 or 1951, in another dataset. Australia has good coverage for the late 1940s – on in another dataset.

Pibals for Reanalysis

(Ave. Counts per Day for 1942 - 1990)

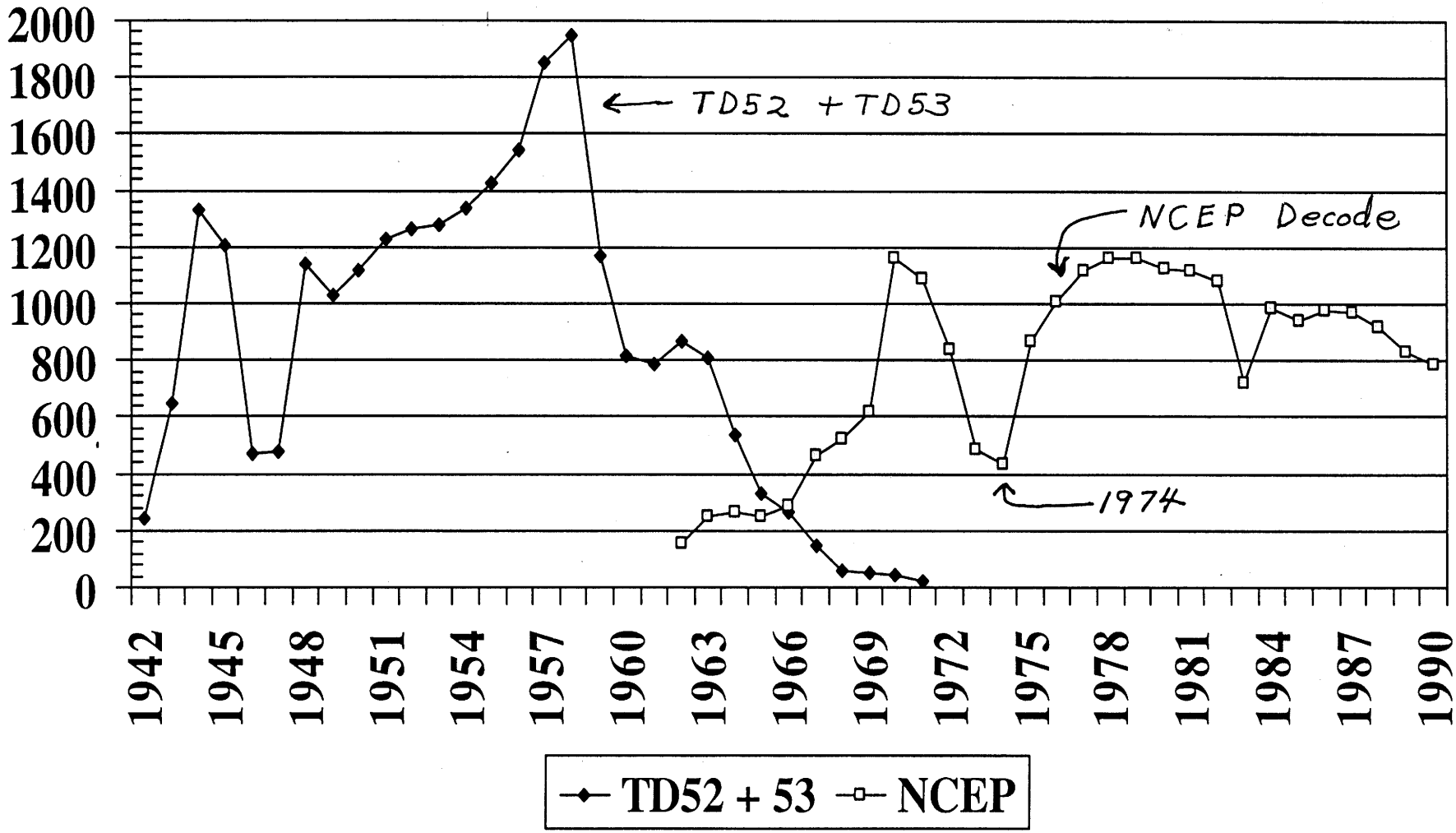
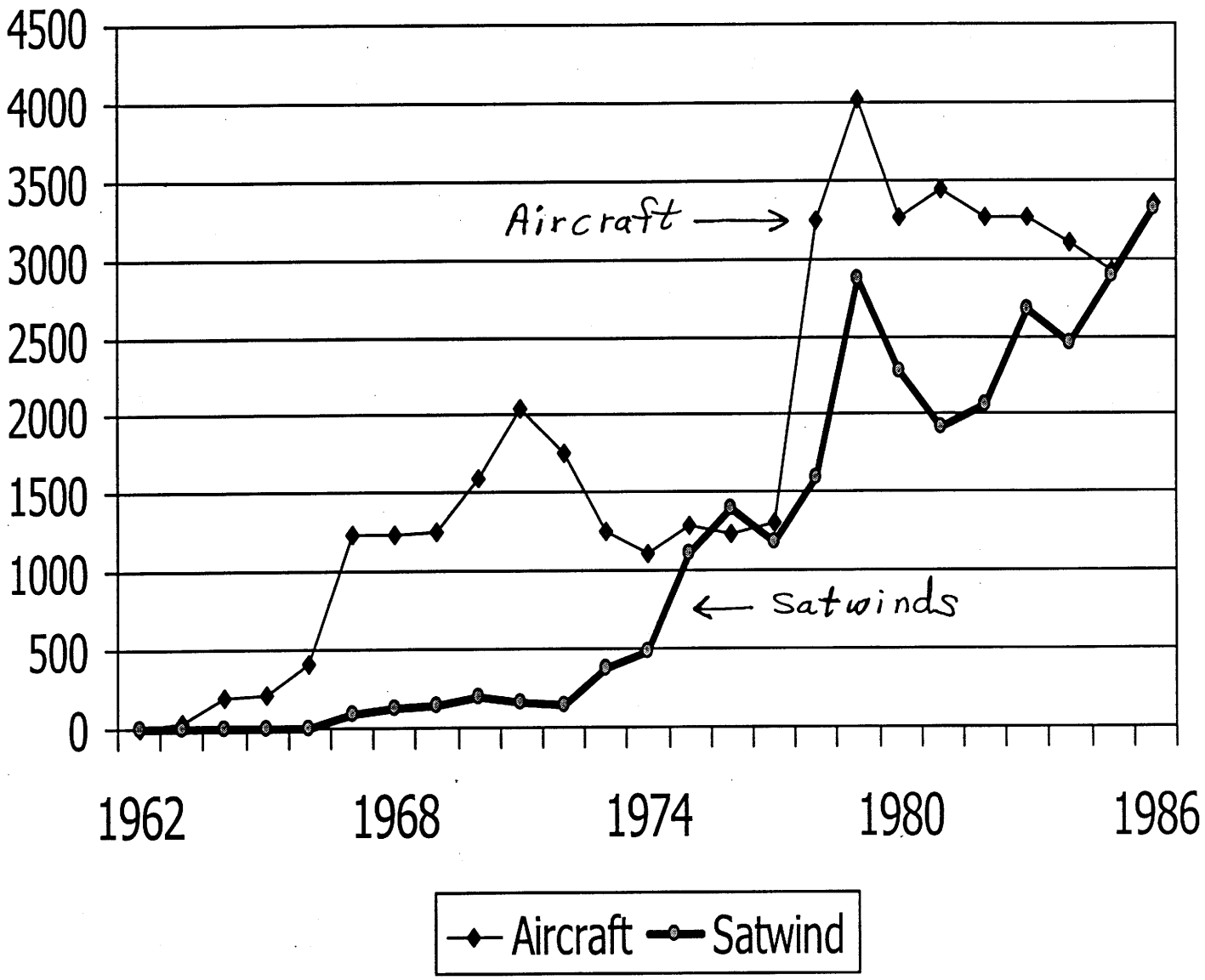


Figure 4. Counts of pibals (wind by height reports) in the main datasets used for reanalysis. Some pibal reports were also received directly from countries.



- Figure 5. The number of aircraft and satwind observations used for the original operational analyses at NCEP are given. Since 1986 the number of aircraft reports has increased to 14,420 in 1998, not counting aircar data. Also the number of satwind reports increased to 18,368 per day in 1998.

Roy Jenne
April 1999

Satellite Cloud Wind Data
(10-day counts in 20-degree long. bands)

Table 1 TEN-DAY COUNTS OF CLOUD WINDS, 1973-1998 (NCEP real-time tapes)

	0E - 20E	20E - 40E	40E - 60E	60E - 80E	80E - 100E	100E - 120E	120E - 140E	140E - 160E	160E - 180E	180W - 160W	160W - 140W	140W - 120W	120W - 100W	100W - 80W	80W - 60W	60W - 40W	40W - 20W	20W - 0W	
1973Q1	3	67	2	3	9	7	80	7	7	25	23	69	391	630	383	543	446	3	
1973Q2	0	46	9	26	26	14	66	16	15	34	22	63	418	843	615	929	855	0	1973
1973Q3	0	36	9	26	37	28	62	30	23	24	7	60	471	951	643	896	869	0	
1973Q4	0	25	2	14	19	15	49	13	10	23	4	52	518	1079	658	934	875	0	
1974Q1	0	36	1	2	9	11	76	8	5	20	8	53	450	856	594	956	995	1	
1974Q2	0	39	6	17	19	12	59	21	18	30	9	54	449	860	760	1089	1076	3	
1974Q3	1	25	0	0	0	0	0	0	0	0	0	42	396	801	769	1164	1293	259	
1974Q4	1	18	0	0	0	0	0	0	0	0	0	121	950	1214	1037	1322	1149	184	
1975Q1	0	0	0	0	0	0	0	0	0	0	0	366	1808	2046	1479	1850	1148	0	
1975Q2	0	0	0	0	0	0	0	0	0	105	1302	1668	1617	1723	1214	1557	887	0	
1975Q3	0	0	0	0	0	0	0	0	0	301	2408	2637	2292	2187	1499	1648	1047	0	
1975Q4	0	0	0	0	0	0	0	0	19	553	2461	2992	2279	2144	1770	1959	1128	0	
1976Q1	0	0	0	0	0	0	0	0	265	1831	2508	2716	2023	2040	1481	1886	1216	0	
1976Q2	0	0	0	0	0	0	0	0	281	2064	2686	2577	1689	1714	1321	1753	1351	0	
1976Q3	0	0	0	0	0	0	0	0	261	1726	2205	2146	1443	1591	1192	1522	1115	1	
1976Q4	0	0	0	0	0	0	0	0	216	1316	1807	1785	1401	1561	1099	1348	1083	4	
1977Q1	0	0	0	0	0	0	0	0	242	1329	1810	1808	1350	1548	951	1266	985	2	
1977Q2	0	0	0	0	0	0	0	0	238	1412	1892	1871	1253	1346	953	1317	907	0	
1977Q3	0	0	0	0	0	0	0	0	267	1468	1929	1805	1290	1500	982	1269	1045	0	
1977Q4	0	0	0	0	0	0	0	0	312	1545	2182	2202	1595	1685	1073	1433	1091	0	
1978Q1	0	0	0	0	0	0	0	0	359	1757	2234	2412	1881	2129	1200	1726	1382	1	
1978Q2	0	0	0	0	0	0	0	0	438	2076	2645	2550	1764	1965	1120	1739	1362	0	
1978Q3	0	0	0	0	0	0	0	0	480	2293	2879	2545	1840	2174	1260	1808	1482	0	
1978Q4	0	0	0	0	1	111	180	253	706	2058	2598	2640	1840	1976	1120	1593	1269	2	
1979Q1	1812	1181	183	0	14	639	1277	1691	1908	2302	2284	2460	1632	1641	975	1941	3203	2022	
1979Q2	1601	836	196	0	15	753	1565	2031	2125	2445	2484	2271	1447	1663	964	1797	2909	2071	1979
1979Q3	2013	1078	229	0	23	743	1483	1944	2264	2935	2774	2458	1684	1750	1073	2079	4063	3137	
1979Q4	1258	847	89	0	19	733	1355	1912	2106	2843	2816	2878	1931	2104	1251	1999	3075	1679	
1980Q1	0	0	0	0	16	735	1314	1900	2168	2616	2521	2670	1949	2135	1276	1881	1545	20	
1980Q2	0	0	0	0	19	738	1529	1899	2279	2941	2907	2702	1838	1999	1295	1925	1650	25	
1980Q3	0	0	0	0	21	688	1299	1754	2193	2965	2853	2441	1850	1948	1408	1971	1526	14	
1980Q4	0	0	0	0	17	671	1210	1556	2041	2607	2584	2581	1889	1842	1341	1842	1489	26	
1981Q1	0	0	0	0	18	714	1303	1968	2064	2181	1988	2129	1718	1840	1045	1536	1297	13	
1981Q2	0	0	0	0	31	776	1471	1893	2171	2363	2208	2086	1380	1519	816	1257	1220	6	
1981Q3	0	0	0	0	37	762	1358	1686	2050	2750	2452	2091	1408	1514	823	1338	1297	3	
1981Q4	0	0	0	0	29	660	1181	1476	1815	2336	2151	2149	1584	1710	829	1321	1100	6	
1982Q1	0	0	0	2	52	754	1297	1850	1989	2167	1832	2071	1555	1705	753	1251	1167	3	
1982Q2	0	0	0	0	36	980	1361	1804	2218	2273	1953	1850	1200	1273	614	1178	1168	4	
1982Q3	486	354	66	0	233	1063	1377	1876	2383	2448	2216	1974	1263	1436	770	1500	2021	624	
1982Q4	1459	1106	198	1	362	1076	1208	1690	2119	1690	1317	1367	1325	1615	805	1457	3117	2111	
1983Q1	1557	1115	189	0	346	986	1176	1777	2062	1535	1102	1119	1119	1441	792	1687	2723	2222	
1983Q2	1760	1053	215	0	356	1081	1471	1969	2273	1924	1402	1328	762	970	626	1554	3403	2935	1983
1983Q3	1472	797	236	0	415	1241	1532	2022	2578	2948	2434	2100	1310	1325	802	1731	3813	2866	
1983Q4	1655	1121	119	0	572	1546	1834	2395	3168	2646	1756	1846	1247	1337	643	1427	3319	2609	
1984Q1	1658	1000	195	0	396	1133	1480	2136	2610	2280	1371	1456	1291	1401	597	1439	2963	2189	
1984Q2	1905	1097	410	0	248	840	1250	1730	2263	2549	1887	1621	1010	1089	581	1901	4087	3015	
1984Q3	1429	887	340	0	23	78	92	86	264	991	1729	1520	1047	916	475	1063	2760	2403	
1984Q4	2441	1787	524	44	526	1764	1679	2282	2599	1286	1375	1711	1315	1218	526	967	3641	3674	
1985Q1	2463	1831	559	181	200	1449	1702	2469	2663	1686	1309	1439	1097	1129	474	1178	3053	2858	
1985Q2	2158	1349	484	265	216	1390	1826	2410	2801	1869	1470	1495	1056	1079	478	920	3051	3219	
1985Q3	2024	1210	451	350	225	1613	1903	2517	2846	1870	1572	1660	1370	1197	596	882	2995	3152	
1985Q4	2173	1603	473	304	369	1931	1861	2386	2662	2094	2066	2199	1757	1577	728	1009	3236	3209	
1986Q1	2112	1441	386	310	238	1446	1742	2429	2403	1977	2313	2268	1713	1546	680	853	2707	2795	
1986Q2	2500	1464	929	359	209	1447	1873	2441	2736	2001	2277	2164	1539	1503	628	1058	3447	3738	
1986Q3	2710	1452	560	536	382	1575	1768	2426	2718	1713	1844	2201	1655	1521	959	1551	4062	4597	
1986Q4	2987	2229	695	557	495	1692	1654	2236	2418	1683	1930	2270	1901	1765	1051	1909	4611	4475	
1987Q1	3369	2307	658	822	561	1314	1595	2198	2482	1797	2106	2334	1852	1816	975	1982	4196	4468	
1987Q2	3308	2285	1042	794	757	1517	1876	2744	3530	2886	2298	2453	1746	1797	1029	2569	5427	5192	
1987Q3	3083	1987	1074	1166	1265	1638	2053	2834	3493	3866	3313	2997	2656	2710	1707	3212	6334	5206	
1987Q4	3274	2472	987	1315	1787	2628	2606	3437	3937	3599	3166	2990	2539	2598	1908	3678	6559	4624	
1988Q1	3237	2402	842	619	1325	2252	3009	4495	4722	3627	2733	2700	2264	2318	1546	3378	5719	4404	
1988Q2	3276	1967	899	621	1407	2624	3324	4606	5143	4324	2919	2772	1943	2281	1580	3509	6177	5195	1988
1988Q3	3043	1687	1142	898	1523	2941	3898	4905	5782	4657	3084	2597	2367	2794	2009	3351	6819	5353	
1988Q4	3030	2198	776	686	1748	3209	3950	4612	5101	4562	3723	3402	2591	3372	2544	4089	6874	4746	

Europe Japan US GOES EU

Figure 6. Winds derived from cloud motion in satellite images. These counts are of data used in NCEP operations. Another text at NCAR has counts through 1998. Reanalysis also used the FGGE datasets (1979) and that includes coverage of cloud winds over the Indian Ocean. We hope to improve data coverage from Meteosat in the early 1980s, for the Europe – Africa sector.

SURFACE SYNOPTIC DATA

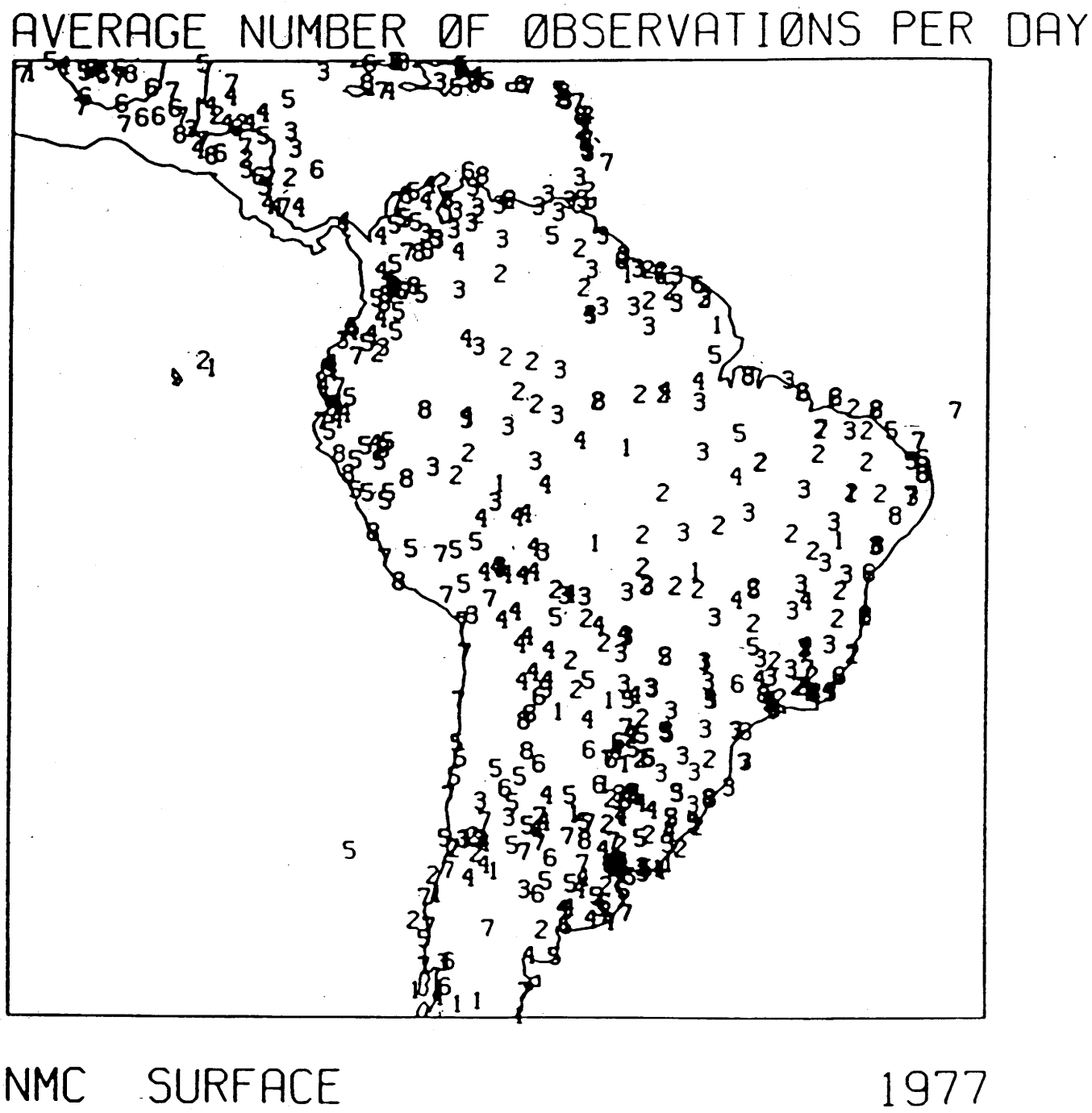


Figure 7. Coverage of surface synoptic observations for Latin America for 1977. The data are from NCEP (was called NMC) GTS tapes. A plotted number 5 means that there is an average of 5 reports per day during 1977. Reanalysis has GTS data from 1967 – on. We think that the data coverage for 1967 – on will be similar to this.

COVERAGE OF NCEP SURFACE SYNOPTIC DATA

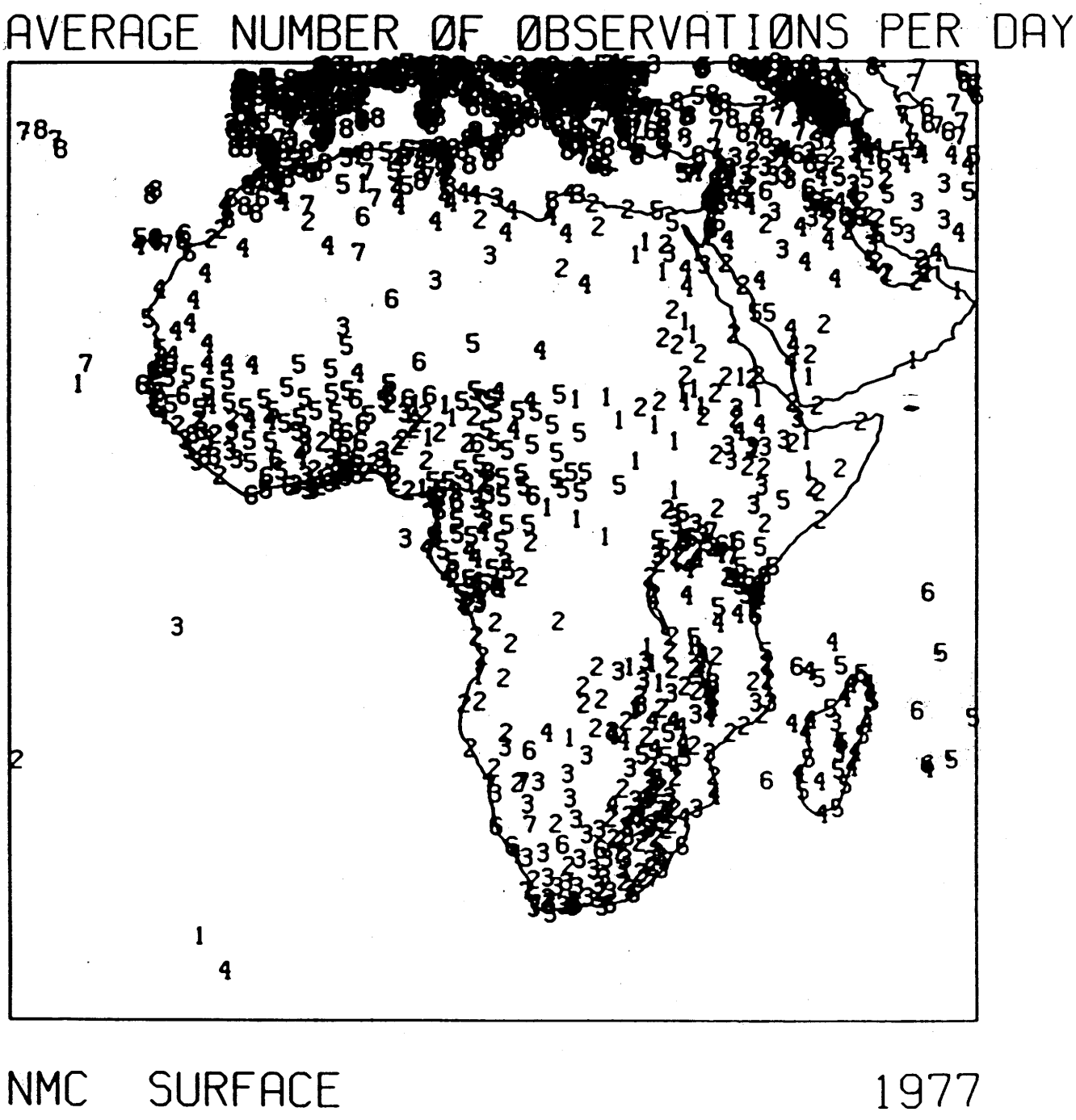


Figure 8. Coverage of surface synoptic data observations for Africa for 1977. The data are from the NCEP GTS tapes. Reanalysis has GTS data or 1967 – on. We think that the data coverage for 1967 – on will be similar to this.

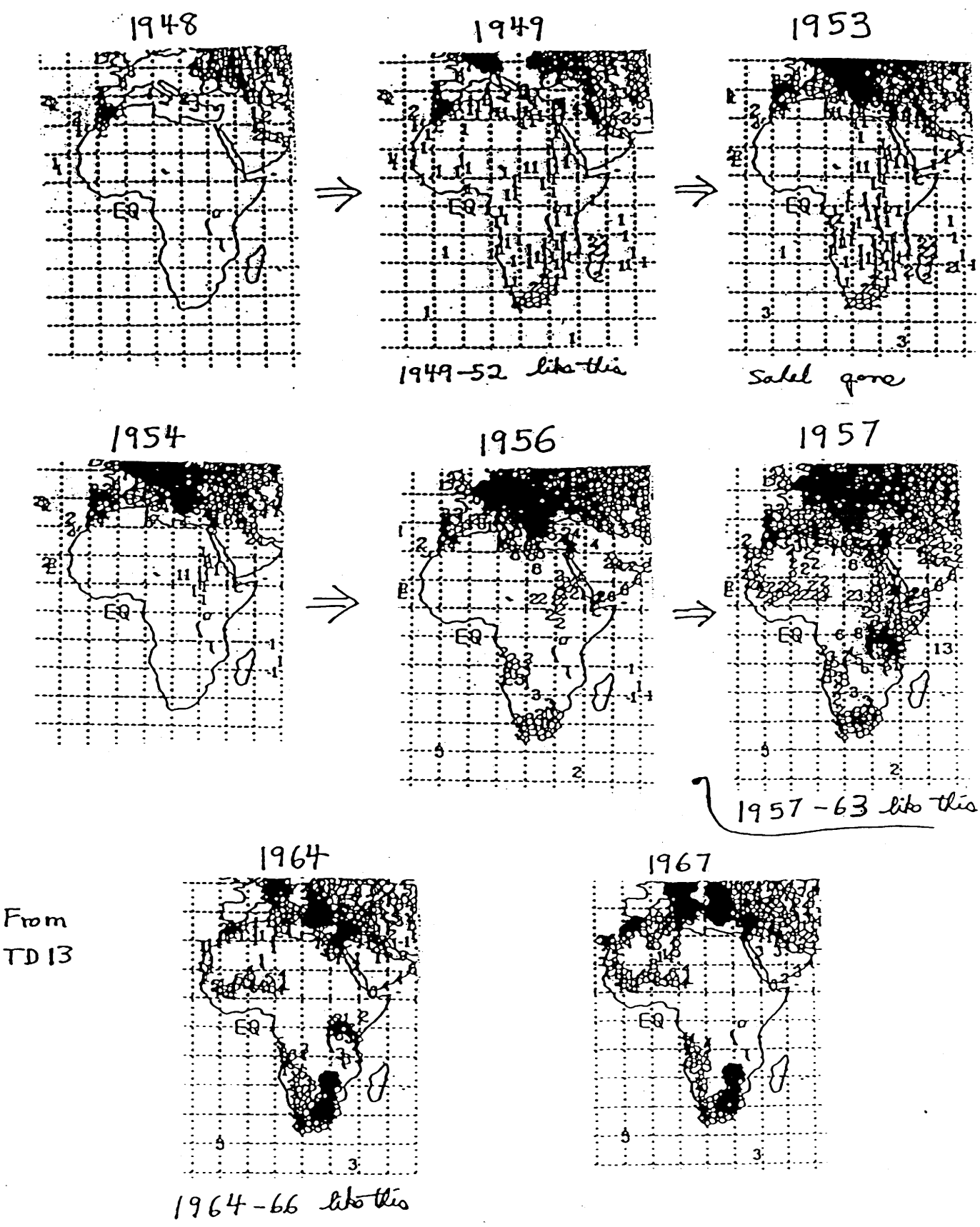


Figure 9. Surface synoptic data coverage for Africa from dataset TD13. The plotted numbers are the average number of observations per day.

SURFACE SYNOPTIC DATA COVERAGE IN TD13

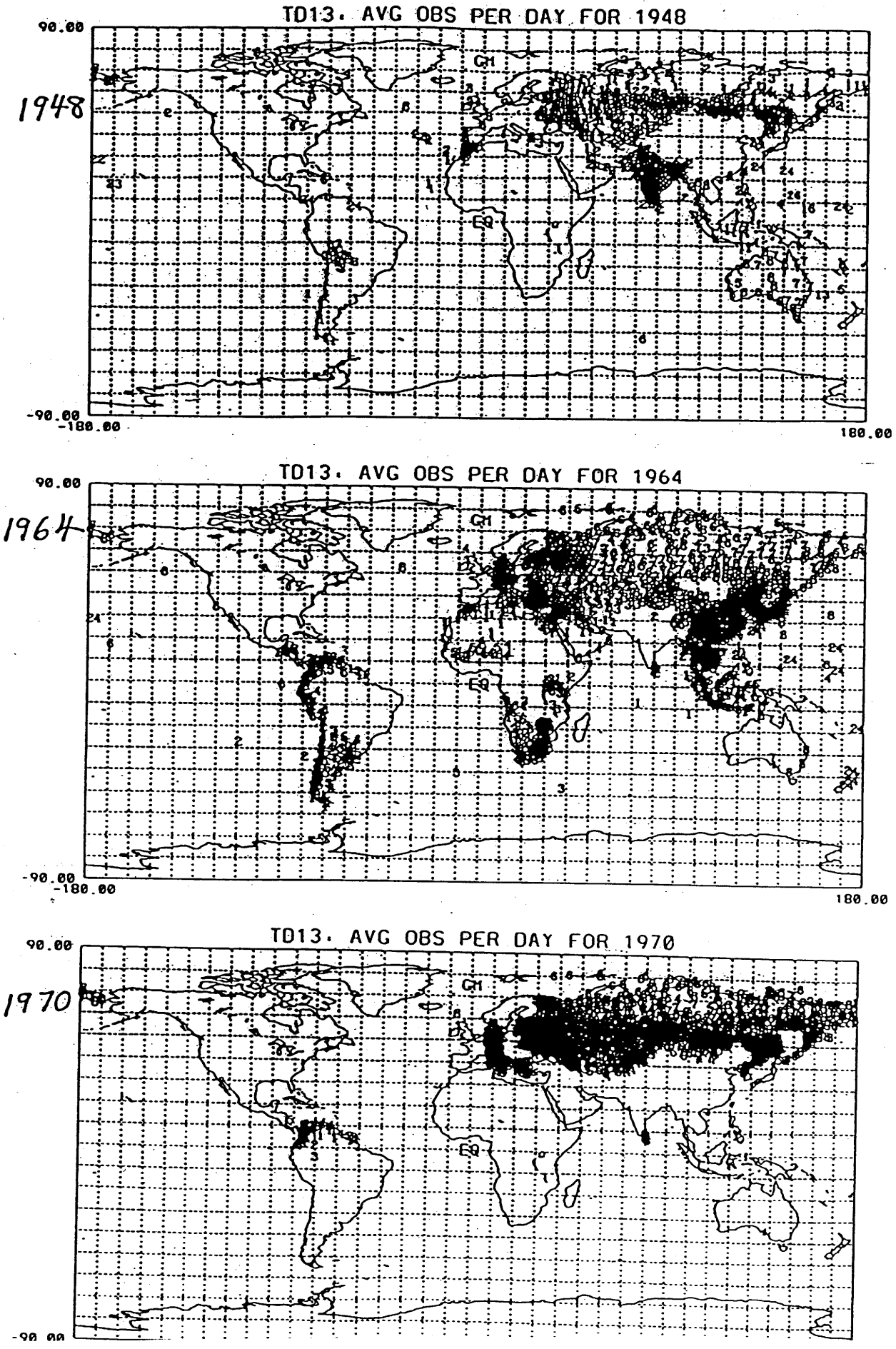


Figure 10. World surface synoptic data coverage from dataset TD13 for selected years: 1948, 1964, 1970.

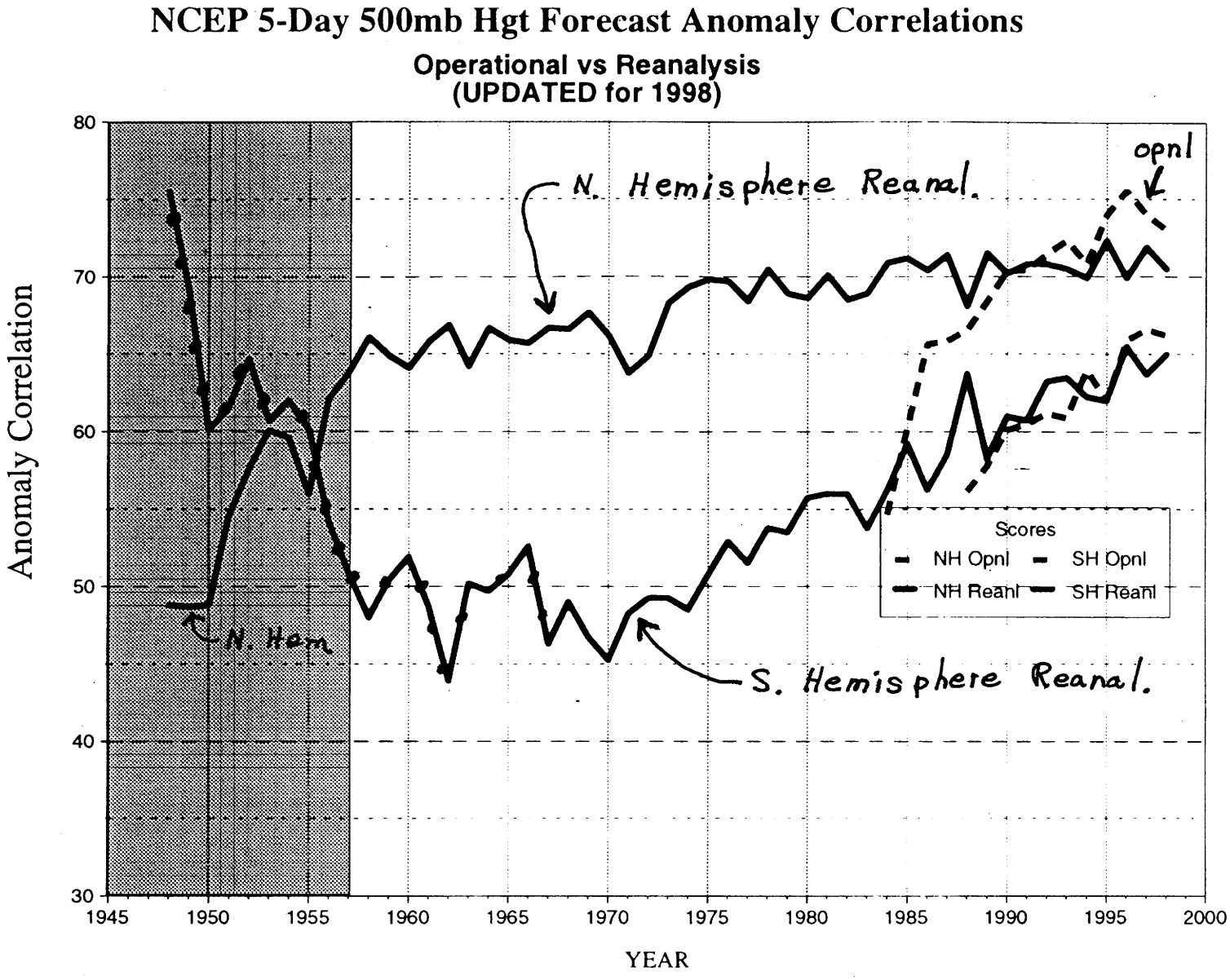


Figure 11. NCEP five-day forecast scores for reanalysis are given in heavy lines for each hemisphere. We think that the drop in 1971 – 72 (N. Hemisphere) was caused by a drop in observations. The forecast scores for reanalysis, for 1952 – 1998 in the N. Hemisphere are better than the operational scores in 1984. This chart is from Kistler and Kalnay in May 1999. An earlier 1997 version had a mechanical problem.