DOE/ER/61011--T5

A Comprehensive Aerological Reference
Data Set (CARDS)

by

Robert E. Eskridge, Arthur C. Polansky, Oleg A. Alduchov¹,
Stephen R. Doty, Helen V. Frederick, Irina V. Tchernykh¹, and
Zhai Panmao²

National Climatic Data Center
National Environmental Satellite, Data, and Information Service
National Oceanic and Atmospheric Administration
Asheville, NC 28801

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

1. Permanent address: Russian Research Institute of Hydrometeorological Information, 6, Korolyov St., Obninsk, Kaluga Reg., 249020, Russia.
2. Permanent address: National Meteorological Center, 46 Baishiqiao Rd, Beijing, P. R. China

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

ABSTRACT

The possibility of anthropogenic climate change has reached the attention of Government officials and researchers. However, one cannot study climate change without climate data. The CARDS project will produce high-quality upper-air data for the research community and for policy-makers. We intend to produce a dataset which is: easy to use, as complete as possible, as free of random errors as possible. We will also attempt to identify biases and remove them whenever possible.

In this report, we relate progress toward our goal.

We created a robust new format for archiving upper-air data, and designed a relational database structure to hold them. We have converted 13 datasets to the new format and have archived over 10,000,000 individual soundings from 10 separate data sources. We produce and archive a metadata summary of each sounding we load. We have researched station histories, and have built a preliminary upper-air station history database. We have converted station-sorted data from our primary database into synoptic-sorted data in a parallel database.

We have tested and will soon implement an advanced quality-control procedure, capable of detecting and often repairing errors in geopotential height, temperature, humidity, and wind. This unique quality-control method uses simultaneous vertical, horizontal, and temporal checks of several meteorological variables. It can detect errors other methods cannot.

We have supported research into the statistical detection of sudden changes in time-series data. The resulting statistical technique has detected a known humidity bias in the United States' data. We expect to detect unknown changes in instrumentation, station location, and data-reduction techniques with this method. We have received software which corrects radiosonde temperatures, using a physical model of the temperature sensor and its changing environment. We need an algorithm for determining cloud cover for this physical model; we have a promising lead. We have a numerical check for the station elevation which has identified documented and undocumented station moves. We have made considerable progress toward algorithms to eliminate one known bias.

We are on track to produce a 5-year quality-controlled subset of the CARDS dataset by the end of the year. The more-difficult problem of bias detection and elimination will take longer. The resulting dataset will justify the delay.

I. Introduction

Current numerical climate models predict strong surface warming in the polar regions and marked cooling in the stratosphere when the model atmospheric CO₂ concentration is doubled. Verification or refutation of these model results requires the existence of reliable long-term data records. The first objective of the USA's Global Change Research Program is:

... long-term records derived from frequent and well documented global measurements of environmentally important parameters are critically needed. Global measurements from satellites and surface-based networks are crucial.

In 1987, NOAA produced the Comprehensive Ocean-Atmosphere Data Set (COADS), containing surface data (Woodruff et al, 1987). A similar data set is needed for the upper air data. The goal of the CARDS project is to produce an upper air data set based on radiosonde and pibal observations, suitable for evaluating climate models and detecting global change. The CARDS project will:

- 1) Produce a long-term (1950-1990) daily dataset of concomitant upper-air and surface synoptic observations using the entire global collection of upper-air and co-located surface observations;
- 2) Develop algorithms to correct and flag rough errors: transmission and keying errors;
- 3) Assess the homogeneity of the data from the upper-air network and implement corrections for biases where appropriate;
- 4) Analyze these data using basic climate change analysis schemes to help ensure against undetected errors and biases;
- 5) Develop software for the operational ingest of future data into the CARDS database; and
- 6) Make these datasets readily available to the research community through NCDC/World Data Center-A (WDC-A), and other institutions.

Meteorological observations of the troposphere and stratosphere have been used to monitor and understand climate variations using two different approaches. Radiosonde data have been used extensively to produce time series of zonal and global temperature and moisture changes (Angell, 1988; Angell, 1986; Angell and Korshover, 1984; Angell and Korshover, 1983; Angell and Korshover, 1978b; Angell and Korshover, 1978a; Angell and Korshover, 1975; Elliott, 1989; Karoly, 1989; and Oort and Liu, 1992). In addition, a number of researchers have used

operational initializations from numerical models to understand long-term climate variations (Knox, et al., 1988).

The advantages and disadvantages of each approach are discussed by Knox, et al. (1988) and by Trenberth and Olson (1988). Radiosonde data are sparse over the oceans, rendering any analysis difficult. Trenberth (1989) has shown radiosonde data can produce misleading analyses even over data-rich mid-latitude continents. Spatial analyses of the data from the radiosonde network can produce spurious results. For example, empirical orthogonal functions derived from radiosonde data may represent the geographic locations of the radiosonde launch sites as well as the condition of the atmosphere. Interpolation procedures are likely to suffer. Operational analyses incorporate a significant amount of supplementary data, including satellite retrievals, manually inserted (bogused) data, and first-guess forecast fields generated by dynamically consistent models. Operational analyses are less sensitive to country-wide biases in radiosonde data. Comparisons between data and the first-guess field can detect many of the regional problems. Data flagged for rough errors (transmission errors) can be corrected or rejected. For example, Böttger et al., (1987) used first-guess analyses from the European Center for Medium-Range Weather Forecasts (ECMWF) to identify numerous errors and biases in existing radiosonde data. On the other hand, changes in the operational analyses can produce biases in the assessment of long-term climate variations (Parker, 1980; Trenberth and Olson, 1988). The "nudging" of data toward the first-guess can suppress real features in data-sparse regions and real features below the effective spatial or temporal resolution of the model. Furthermore, changes in satellite algorithms and bogusing techniques will affect these fields. The correction of biases in radiosonde data fields through the use of operational model initializations is a very difficult problem.

The ECMWF analyses as well as the NMC analyses (Julian, 1989) have allowed researchers to identify many errors and biases in the existing radiosonde network. For example, a systematic, diurnally varying radiation-induced error of about 30 gpm in the 100 hPa geopotential fields exists in the North American radiosonde data (Böttger et al., 1987). Radiosondes in all of India have been consistently put on a suspect list by national centers. Radiosondes used in many other parts of the world have an assortment of biases, many of which have been documented over the past few years by both NMC and ECMWF in internal publications.

International radiosonde comparison experiments are another means of estimating biases in radiosonde measurements. Unfortunately, comprehensive standardization of all the biases would require comparisons of the full suite of radiosonde types ever used. This would not be feasible, even if all instrument types were still available. Since radiosonde biases are time and location dependent, tests would have to be carried out in many portions of the world and at several times of day. Nonetheless,

the studies conducted by Nash (1984), Nash and Schmidlin (1987), and Ivanov, et al. (1991) are very useful. These studies show the likely magnitudes of the errors, and how the biases vary with the time response of the instruments and with the physical properties of the radiosonde. For example, Nash and Schmidlin (1987) found a -40 gpm bias at 100 mb (Hpa) in Australian radiosondes. Other radiosondes had 100 Hpa biases in the range of -40 to 40 gpm. A difference of 65 gpm at 100 Hpa corresponds to a layer mean temperature error of 1° C. Humidity errors were also found. They were a function of the response of the sensor. Errors were least for carbon hygristors (2 to 5%) and greatest for goldbeater's skin sensors (10 to 12% errors).

Atmospheric humidity measurements are particularly sensitive to changes in instrumentation and procedures. It has been difficult to design and produce an inexpensive humidity sensor capable of responding quickly to the large vertical variability of humidity in the atmosphere. The problem is still not entirely solved. The following history of the United States' operation of radiosondes show how often systematic changes are introduced into the data.

1938	Hair hygrometers and bimetallic strip thermometers were used. Balloons were launched at 4 and 16 GMT
1943	Hair hygrometers and lithium chloride instruments were used.
1946	A change was made from a constant height to a constant pressure analysis.
1948	Humidity was calculated with respect to saturation of water instead of ice for temperatures below freezing. Previously, saturation with respect to ice was used.
1948	Observation times were changed from 04 and 16 GMT to 03 and 15 GMT.
1950	A radiation correction to the temperature data was introduced.
1957	Time of radiosonde launching was changed from 03 and 15 GMT to 00 and 12 GMT.
1960	The radiation corrections of temperature data were discontinued.
1961-1965	The lithium chloride hygristor was replaced by a carbon hygristor. A defective housing allowed solar radiation to warm the hygristor, causing systematic low values of relative humidity.

- 1965 US began reporting all humidity measurements. The earlier practice was to censor (not report) humidity values when the temperature was below -40C.
- 1969 The manual data reduction system was replaced by a computerized system.
- 1970-1971 A temporary "quick fix" kit to repair the defective housing was distributed.
- 1971-1973 A new housing was introduced to fix the dry bias of the carbon hygristors.
- 1973 US stations began censoring low relative humidities. All relative humidities below 20% were reported as 19%, or as dewpoint depression 30C.
- 1981 A modified carbon hygristor was introduced.
- 1988 The resistor in the humidity circuitry was changed, however the National Weather Service failed to change the data reduction program until 1993.
- 1993 All relative humidity values will be reported and the value of g is being changed to the WMO standard value.

These operational changes introduced systematic errors in all of the measured variables, and introduced inhomogeneities into the long term record.

Gaffen (1992) has conducted a survey of all historical radiosonde instruments for each of the countries in the World Meteorological Organization. This work has been partially supported by the CARDS project. The results of this survey should prove valuable in assessing potential biases in the global radiosonde network.

We cannot see any scientific justification for ignoring the potential biases and inhomogeneities in the upper-air data record. Yet, even today, research studies of global tropospheric and stratospheric variability do not address these problems. For example, Angell (1988) used a 63-station network of radiosondes to monitor layer-mean temperatures in various portions of the globe. He did not address the homogeneity of his dataset. Oort and Liu (1992) used a network of 700 to 800 radiosonde stations for the period 1958-1989. They found "excellent agreement with ... Angell's 63-station network results". Oort and Liu conclude that the worldwide radiosonde network provides adequate information to estimate some of the most important areal-mean climate parameters, including mean hemispheric temperatures and

long-term temperature trends. However, they used a set of stations with potential uncorrected biases similar to those in the Angell network. While we believe our criticisms of these studies are valid, we understand why the necessary bias removal was not done. Quality control and bias removal are difficult, tedious, time-consuming efforts. Investigators must make do with the information at hand, if they wish to obtain any results at all. The lack of a readily-available, comprehensive, quality-controlled, homogeneous, and unbiased upper-air dataset has introduced potential errors in these competently performed studies. The CARDS dataset will reduce this source of error by providing ready access to a high-quality upper-air dataset.

II. A brief history of upper air measurements

Upper-air measurements were taken as early as the 1620s by Kepler, who measured cloud height by using the cloud shadow on the surface. The development of the balloon by Bartholomeu Loarenco de Gusmao of Portugal in the early 1700s (Americana, 1987) led to the development of methods for measuring pressure, temperature and humidity in the atmosphere above the earth's surface.

The earliest known measurement of the vertical structure of the atmosphere was made by the French scientist Perviet, in 1647. He carried a mercury barometer up Mt. Pew de Dom, and discovered that atmospheric pressure decreases with height.

The invention of the manned balloon in 1783 in France (Middleton, 1969) and/or 1731 in Russia by Kryakutny (Zaitseva, 1990) facilitated the measurement of pressure, temperature, and wind at various elevations. The measurement of winds at night and in the absence of clouds were possible with manned balloons.

During the period from 1860 to 1900 many manned balloon flights were undertaken for scientific and military purposes. Series of flights were undertaken especially in England, Germany, and Russia (Zaitseva, 1990). Zaitseva (1990) gives an interesting and detailed account of manned flights by Russians. An equally interesting account of early ballooning is found in Turnor (1865).

Wind and/or cloud speeds were first measured in 1766 by Alexander Brice of Scotland. Special instruments for observing cloud heights and motions begin to appear in the 1800s. Instruments developed include: horizontal reflector, mirror nephoscope, direct-vision nephoscope, ceiling projector, and range finder. The design and use of these instruments can be found in Middleton (1969).

Strong winds and large wind shears can destroy a manned balloon. Hence, to ensure the safety of the crew, detailed wind information was needed before the balloon flight. Small unmanned balloons called "pilot" balloons were often launched before the launch of manned balloons, to gauge the wind. On December 1, 1883, the date of the first manned hydrogen balloon ascent, a

pilot balloon was released just before the manned flight. The first known pilot balloon observation taken independently of manned balloon flights was taken by Wallis in 1809 (Middleton, 1969).

Cleveland Abbe in 1888 (Middleton, 1969) proposed making regular launches of pilot balloons and transmitting the observed wind data. Abbe proposed to the US Signal Corps the following suggestion:

"This method, which is not expensive, is worthy of being made a special feature of all first-class meteorological stations and national services. The flight of such a balloon during five or ten minutes would be an item more important for the daily weather map than the usual local record of wind, and, as recommend by me in 1872, should be added to the ordinary weather telegram".

Abbe recommends:

"The balloon should carry a suspended light thread from 50 to 500 feet long, at the bottom of which hangs suspended a light object. The observer can at any time ascertain the linear distance and altitude by observing the apparent angular altitude of the upper and lower end of the vertical line thus carried by the balloon..."

In 1903 Hergesel demonstrated that the rate of rise of a balloon depends only on its weight and the lifting force. Hence its position was a function only of time. In 1904 De Querrain designed an optical theodolite, whose basic design is still used today. Numerous improvements and modifications were made to the theodolite between 1905 and 1920.

The first balloon experiments with instruments able to automatically record temperature and pressure were performed in 1892 in Paris. After the balloon burst the instrument package was returned to the surface by parachute. In 1902 Teuceiran de Bord discovered the stratosphere using these balloon-instrument packages.

The United States has operated an upper air observation program since 1907, when a kite station was opened. Pilot balloon observations were started in 1918 with the establishment of five stations. The network grew and reached its peak in 1944 with 147 stations (see Fig. 1). The most important limitation to pilot balloons is that one must maintain visual contact. In 1923 the U.S. Army Signal Corp began experimenting with a buzzer radio transmitter which was tracked for up to 20 minutes. In 1928 the Signal Corp used a vacuum tube transmitter with some success. Optical tracking agreed within one degree of the electronic tracking.

The radiosonde was developed independently in France, Germany, and the Former Soviet Union (FSU) in the 1920's. The first successful launch of a radiosonde took place in France on March 27, 1927 by P. Idrac and R. Bureau (Middleton, 1969).

The first operational radiosonde launch took place in the FSU (Zaitseva, 1990) on January 30, 1930 from the Pavlovsk Aerological Observatory. The data from the sounding were

transmitted to the Leningrad Weather Forecast Bureau. The FSU's radiosonde network was established in 1937 and grew to 220 stations in the 1980s. The breakup of the USSR has disrupted the network and reduced the number of stations

The first radiosonde station opened in the U.S. in 1937 and the network increased in size to 138 stations by the end of 1960. The network has decreased to 96 National Weather Service operational sites in 1990.

Figures 2 and 3 show the distribution of pibal and radiosonde sites world-wide in 1980 and the number of radiosonde stations from 1970 to 1990.

III. The CARDS project

a. Datasets and decoding

In order to build a world-wide upper-air dataset for the 1950 to 1990 period the CARDS project has selected several datasets for inclusion in the CARDS database (see Table I). Each dataset has a long and varied history. Processing systems, quality control efforts, media changes, and numerous other procedures have introduced a variety of idiosyncrasies and errors into the data. Formats and datasets are seldom documented to the extent needed by the most demanding upper-air data manager. The most perplexing problem facing the CARDS project team on a day to day basis is the conversion of input datasets to a common format. Without a consistently formatted and constructed dataset, identifying the best observation from many duplicate observations is virtually impossible.

We decided that as much of the original data as possible would be preserved in the conversion to the CARDS standard format, identified as TDF63. The TDF63 format allows for an unlimited number of levels, contains a quality flag for every element for each level, and allows entry of edited data while maintaining the original value. The format was designed to be user-friendly on a variety of common computers yet flexible enough to accommodate even the most detailed upper-air observation. The TDF63 format has also been adopted as the official National Climatic Data Center format used in archiving upper air observations.

One of the first datasets to be converted to the new format was NCDC's own U.S. National dataset, TD6201. This dataset contains data from 1948 to 1992. The conversion was reasonably straightforward, with the construction of a station identification table the largest obstacle. The upper-air format requires World Meteorological Organization (WMO) station numbers, latitude, longitude, and elevation, where the U.S. National dataset has a special U.S. code number, the Weather Bureau, Army, Navy (WBAN) number, for station identification. The identification of data levels (flags) caused other problems as mandatory levels were sometimes coded as a tropopause only, and the surface level was miscoded in the period 1981 to 1986. The presence of data levels without data prompted us to insert code

to delete these no-data levels during the preload/quality control process.

Data received from the National Meteorological Center (NMC) were especially difficult to decode and convert to the new format, TDF63. A typical observation's levels were not stored in chronological or pressure order. The mandatory levels, significant thermodynamic levels, wind levels, etc. were kept in separate subsections of the sounding. We found that one level in an NMC sounding might have up to three distinct entries in an observation. We developed code to combine these multi-entry levels into a single level in the new format. Numerous levels were miscoded both as significant and as mandatory levels.

Table I. CARDS data sources.

<u>Source</u>	<u>Coverage</u>	<u>Period</u>	<u>Number</u>		<u>Status</u>
				<u>Obs (M)</u>	
NCDC(TD6201)	U.S.	1946-90	4.8		Loaded
NMC GTS	Global	1973-90	16.0		Partly Loaded
NCAR/NMC GTS	Global	1971-72	0.8		Converted
USSR GTS	Global(1)	1984-90	2.3		Partly Loaded
MIT	Global	1958-63	1.0		Converted
Australian GTS	S. Hemis.	1990-93(2)	0.3		Being loaded
NCDC (TDF56)	Global	1950-70	2.4(3)		Being Converted
NCDC (TDF54)	Global	1945-70	NA		Being Converted
P. R. of China	PRC	1951-90	1.2		Partly Loaded
Argentina	Argentina	1951-90	0.1		Loaded
Netherlands	Netherlands	1945-91		0.04	Loaded
Hong Kong	Hong Kong	1956-90	0.02		Loaded
British Antarctic	Antarctic	1957-91	0.02		Converted (4)
Korea	Korea	1984-92	0.02		Loaded
Hungary	Hungary	1962-90	0.04		Loaded
Australia	Australia	1950-92	NA		Being Converted
USSR	USSR	1960-90	NA		In house
Canada	Canada	Unknown	NA		Requested
Brazil	Brazil	1961-80	0.4		Being converted
NCDC (TD6210)	OSVs	1945-75	NA		Being Converted

1- Contains no USSR stations, 2- Earlier years have been requested, 3-Does not include 1963-70 totals, 4-New source data coming, Converted = source data have been converted to the CARDS standard format, TDF63, Loaded = data have been converted, quality controlled and entered into the Empress database on the optical jukebox. NA = Not Available

CARDS also received NMC data for the period 1971 to 1972 from the National Center for Atmospheric Research (NCAR). These data were in a packed 60 bit binary form. Several major problems arose in the conversion effort, which was difficult because of the packed binary. Observations with the time 23 GMT were assigned to the wrong day. Several times, global data for an

entire synoptic hour (00 or 12 GMT) was repeated in the dataset. For example, the 12 GMT observations for August 22, were repeated at 12 GMT on August 23, with the date "August 23" stored in the observation. The valid 12 GMT, August 23, data were mixed with the duplicates in the data stream. We also encountered:

- invalid station numbers

- invalid station locations (latitude/longitude)

- observations with no data or nothing but illegal values.

Details of the problems encountered in decoding the National Center for Atmospheric Research (NCAR) and MIT datasets are given in the appendix.

A dataset (1984-90) of world-wide radiosonde data, furnished by the Russian Institute of Hydrometeorological Information, were also converted to the new format. A level type code missing from the documentation caused many levels to be initially coded as "unknown level type". They were later confirmed to be winds at constant heights. However, the largest idiosyncrasy was the duplication of winds at mandatory levels. A complete observation including all thermodynamic data and, most of the time, winds was often followed by one of the mandatory level winds-only observations. Again, new software was created to eliminate this duplication.

To date, some 13 different datasets have been converted to the TDF63 format. Each has had its special nuances. However, many of the problems that have been found fall into the nuisance category: getting translations of a non-English reference manual; acquiring adequate definitions of the flagging and coding conventions; having tapes delivered that weren't crushed or even getting them delivered at all. The conversion effort has been one of discovery and continual software revisions.

To ensure that data can be recovered in case of decoding or processing mishaps, data from each phase of the process are maintained in the NCDC official tape library. Thus, the original dataset as received and the original TDF63 formatted version are available. Copies of each subsequent process (quality control, duplication elimination, etc) are maintained on 8 mm tape.

b. Quality control of upper air data

The presence of errors in meteorological data must be taken into account before the data are used. Removing or detecting errors is especially important in any climate change analysis, since the noise (errors) in the observation network and meteorological observations may be larger than the quantity (eg temperature change by decade) being investigated.

The three main types of errors in radiosonde data are: random observational errors, systematic observational errors, and gross (rough) errors. Observation errors are due to inaccuracies in the measurement of atmospheric variables such as temperature, relative humidity, and pressure. The number and statistical structure of observation errors are determined by the quality of the observations. It is not possible to remove observation

errors. But observation errors generally have constant statistical properties and one can take these errors into account by studying their structure.

It is important to differentiate between random and systematic observation errors (Hawson 1970, Hooper 1975). Random and systematic errors are differentiated by their mean value, which is zero for random errors and nonzero for systematic errors. The presence of systematic errors is generally attributed to inadequate or erroneous actions in taking the radiosonde observation, to a change in the instrumentation, or a change in the data processing procedure. These actions cause the emergence of systematic errors (not necessarily constant in time and space) in the observational data. The detection and removal of these errors is a complicated, but necessary, step in the analysis of climate change.

Gross (or rough) errors are caused by mistakes or malfunctions at any stage of data processing. Experience suggest that from 5 to 20 percent of upper air observations contain gross errors (Gandin (1988) and Alduchov (1982)). The percentage depends on the part of the world and the time period. The composition, magnitude, and occurrence of particular types of gross errors varies with each dataset. Gross errors can significantly distort the results of any data analysis. Thus a quality control (QC) procedure is a necessary step in meteorological data processing. The QC procedure's main task is to identify and remove gross errors from the data and it clearly must precede any analysis. Unfortunately, observational errors can not be removed by a quality control procedure. See section (c) below.

1. Quality control procedures.

A QC procedure can be logically defined as follows. The variable being controlled is assigned to one of several classes (subsets) into which the set of observations is divided. Usually the data are divided into two classes: a class of correct values, and a class of erroneous values. Errors in the quality control procedure (QCE) occur when the controlled value is assigned to the wrong class. There are two different types of QCEs: a QCE of the 1st type occurs when an erroneous value is assigned to the class of correct values. In a QCE of the 2nd type, a correct value is assigned to the class of erroneous values. It is clear that the occurrence of these errors is highly undesirable, since they may distort any data analysis.

It is not difficult to develop a QC procedure which can minimize the quality control errors of either type. One can apply, for example, a check for physical limits within sufficiently large bounds or gates. This will minimize errors of the 2nd type. The large bounds will guarantee that not a single correct value is taken as erroneous. However, many erroneous values would be taken as correct. To minimize quality control errors of the 1st type it is possible to use the same procedure with very narrow bounds. All erroneous values would be removed,

but many correct values would be misclassified as erroneous.

The main problem in developing a reliable QC procedure is developing methods which will minimize both types of quality control procedure errors. Experience suggests that a simple single criterion QC procedure can not minimize the QCE's with the current level of upper air data redundancy. Thus, more complex methods have to be developed to quality control data from the atmosphere.

2. Comprehensive hydrostatic quality control

For the first phase of quality control, CARDS elected to implement the Comprehensive Hydrostatic Quality Control (CHQC) system that had been running several years at NMC. CHQC was developed by Gandin and Collins (1989) and Collins and Gandin (1990), and was modified for CARDS use by Huang (1992). CHQC uses the hydrostatic relation to identify and flag up to 30 types of errors and can make confident corrections to temperature and geopotential height data for a number of these errors. It error checks and corrects temperature data at both mandatory and significant levels, and geopotential height data at mandatory levels.

All data are processed by CHQC prior to loading. It provides a standard to judge the quality of each data level. The amount of errors detected by CHQC has varied considerably from data source to data source. Table II gives a sample of the results.

Table II. Results of running Comprehensive Hydrostatic Quality Control on CARDS data sources. POR = Period of Record

	NCDC POR	NMC 71-72	NMC 1988	Arg POR	Aust 90-93	USSR 89-90	PRC 80-90	Hung POR
Total obs (millions)	4.8	0.8	0.7	0.2	0.3	0.79	0.29	0.04
% obs. flagged	1.1	18.0	10.1	5.7	16.6	8.8	8.1	0.09
% obs corrected	0.09	3.7	4.1	2.7	6.4	3.1	4.3	0.04

3. Complex quality control.

The idea of a complex (multicomponent) quality control (CQC) of meteorological data was proposed by L. S. Gandin (1969) and developed under his guidance in other studies: Parfiniewicz (1976), Antsipovich (1980), and Alduchov (1983). This was a new and imaginative approach to solving the problem of meteorological

data quality control. Gandin introduced the idea to combine, through a decision making algorithm (DMA), simple quality control methods (CQC components) whose working logic would be similar to those of a human being. This integrated system results in increased sensitivity to errors, improved determination of errors, and superior decision making. The CQC minimizes the number of quality control errors (QCE) of both types without degrading the positive features of each CQC component.

There is little difference between the use of control procedures within the CQC framework and the individual use of each quality control procedure. A criterion, which serves as the basis of a control procedure, is used to check the data. When a suspected value is found, the DMA weighs the analysis of each CQC component, and makes a decision whether the value is correct or erroneous based on a joint analysis of all CQC components. Such a procedure permits the use of significantly smaller bounds.

There are a great variety of errors, and each CQC component has different sensitivities to these errors. Therefore, the most complicated and important task in the construction of the Complex Quality Control (CQC) is the development of the DMA. Given the error analysis of each individual CQC component, the DMA must weigh the data in each case and make a decision.

4. Main principles of the choice of CQC components.

The choice of quality control components to use in the CQC system is of great importance. For upper air data, it is useful to check observations for mutual consistency with bracketing soundings (time consistency), at adjacent heights (vertical consistency), and with the data of neighboring stations (horizontal consistency). Hence, these types of checks must be components of a CQC for upper air data. In the context of climate change analysis, the horizontal check is of particular importance since this check will reveal systematic observation errors at individual upper air stations.

Time, vertical, and horizontal checking are usually based on the interpolation of observational data to the station being checked. A comparison is made between the results of the interpolation and the observed values. The data interpolation method plays a significant role in the quality control of upper air data. There are many mathematical methods used in the interpolation of data. But, optimal interpolation of upper air data is the preferred method for use in quality control procedures (Gandin, 1963). Optimal interpolation allows not only the accurate interpolation of the data, but gives an estimate of the accuracy of the interpolation at each observation point. Error estimates are used in the quality control procedures. Optimal interpolation has another important advantage compared to other interpolation methods. The advantage of optimal interpolation is that statistics of controlled values (first and second moments) over a field, which are needed for optimal interpolation, are already known from historical data.

Therefore, the QC procedures can take into account the historical behavior of the variables being controlled. The more detailed and reliable the statistics used in the interpolation, the more likely is the local "behavior" of the variable to be controlled correctly.

It is very important during the quality control of upper air data to make sure a sounding is internally consistent. The main criterion of consistency for geopotential height, temperature, and pressure is the requirement that the hydrostatic equation be satisfied. The hydrostatic equation is the basis for one of the most effective QC methods for upper air data.

Tests for internal consistency of geopotential heights, temperatures, and winds are provided by checking the data against the geostrophic and thermal wind equations. These tests use optimal differentiation of the geopotential and temperature fields (Gandin and Kagan 1976).

The ability of quality control to detect and to locate errors in the data depends on the skill to create an accurate prediction of the value in question, and the skill to use several independent predictions of the value. The more accurately we can calculate (predict) the value in question, the smaller the errors we can detect. If we have only a single predicted value of an observation, we cannot be sure which is erroneous: the observation, the prediction, or perhaps both. As a rule, to calculate a predicted value for an observation, we must use observations which are questionable. Therefore, it is necessary to have several independent predictions of each observation to accurately locate erroneous observations.

To quality control the CARDS' upper air data, we have developed a complex quality control (CQC) method which allows us to check geopotential height, temperature, wind speed and direction, and humidity at mandatory and significant levels. The following tests are part of the CQC:

- a comparison of observational data at mandatory levels to horizontal optimal interpolation of data from different stations;
- a comparison of observational data at mandatory to vertical-optimal interpolated data;
- a check of consistency of mandatory and significant levels for each profile;
- a check that geopotential height and temperature satisfy the hydrostatic equation at mandatory levels;
- a comparison of geostrophic winds and real winds at mandatory levels;
- a comparison of the thermal wind to the real wind at mandatory levels.

It is important to know how accurately we can recalculate upper air data, and what errors this quality control procedure can detect? Figures 4 through 9 give some answers to this question.

In Figs. 4 through 8, we show several of the CQC components in comparison with the well known check against climatology. The

rms values were calculated using data from a global set of 759 stations from 00 GMT January 15, 1989. The climatology check compares the observation in question with climatic (monthly) mean values. The difference between the correct data and the climatic mean should not be more than 4 to 5 standard deviations. The results shown in Figs. 4 through 8 can be interpreted as a kind of noise. This noise is a background which limits our ability to detect errors in the data.

In Fig. 4, we see that the rms value of observed minus monthly means is almost 150 gpm between 300 and 100 hPa and increases to 400 gpm at 10 hPa. This QC procedure will detect errors in geopotential height that are greater than 0.5 km between 300 to 100 hPa and greater than 1.5 km at 10 hPa. Clearly, this kind of QC test of geopotential height data is not completely satisfactory.

Horizontal optimal interpolation of geopotential heights at mandatory heights from surrounding radiosonde stations allows us to reduce the rms values of the differences between observed and interpolated values to 40-50 gpm between 300 to 100 hPa, and above 100 gpm at 10 hPa. This QC test will detect errors almost four times smaller than the climatic check (see Fig. 4).

However, there are methods with which much smaller errors in geopotential height data can be detected. These tests are: the hydrostatic check; a vertical statistical check which is based on vertical optimal interpolation of geopotential height data from the nearest mandatory levels; and a horizontal check of geopotential height which is based on optimal interpolation of thickness between each pair of mandatory levels using neighboring stations. Fig. 4 shows that the rms error of these methods for geopotential height at mandatory levels is from 6 to 30 gpm between 1000 and 20 hPa. This means that we can detect errors in geopotential height of 30 to 50 gpm at most mandatory levels. At the top and bottom levels the procedure of vertical interpolation changes from interpolation to extrapolation, resulting in larger errors at these levels.

It is very interesting to see that these three methods (hydrostatic, horizontal interpolation, and vertical interpolation) will detect errors which are smaller than the observational errors in geopotential height at most mandatory levels. The observational errors estimates are from Alduchov (1985). This somewhat surprising result is due to the high correlations between observational errors of geopotential height at adjacent mandatory levels. This correlation arises from the procedure used to calculate geopotential heights at upper air stations.

These three methods will not detect all types of errors in the geopotential height data. Hence, to detect some types of errors, we need to employ horizontal optimal interpolation of geopotential height at mandatory levels. The sophistication of these four methods allows us to detect almost all possible errors in geopotential heights, except very small errors of less than 30 to 50 gpm.

It is important to note that these methods are not applied separately (sequentially), but simultaneously. The decision making algorithm uses evidence from all the methods to make a reliable decision on whether an observation is erroneous or correct.

Fig. 5 shows that the difference in rms between the observed values of temperature and the climatic monthly means is about 4 to 5 °C at most mandatory levels. This means that the well known climatic check can detect temperature errors only when they exceed 15 to 20 °C. The other three methods: horizontal optimal interpolation of temperature from neighboring stations, vertical optimal interpolation of temperature from adjacent mandatory levels, and calculations of temperature using the hydrostatic equation reduce the rms of the difference between observed and calculated temperatures to about 2 °C at most mandatory levels (Fig. 5). Therefore, using these checks one can detect temperature errors of 7 to 8 °C.

The vertical interpolation of temperatures from adjacent significant levels have the smallest rms values, which is not unexpected because of the criteria used to pick significant levels. Using this temperature check we can detect errors in temperature of similar magnitude to observational errors (4 to 5 °C). This is a very nice result. Unfortunately, not every temperature observation has adjacent significant temperature levels and the temperature at these significant levels must also be checked. These issues complicate the use of this temperature check. Hence, we can use this check only in conjunction with horizontal optimal interpolation, vertical optimal interpolation, and/or the hydrostatic check. The combination of tests ensures that we are not using an erroneous value to reject a correct temperature observation. It is the sophistication of these four methods, especially when we check the evidence of one method by another method, that enable us to detect errors in temperature greater than 5 °C.

Figs. 6 and 7 show the accuracy of various checks of the wind vector components. We show here the results of zonal and meridional wind components, because they are much easier to work with than speed and direction. However, the decision making algorithm uses wind speed and direction to detect errors.

Figs. 6 and 7 show that the accuracy of these methods is strongly dependent on the method and some of these methods are height dependent. Comparison of the observed winds with climatic data shows that the rms values range from:

4 to 5 m/s at the lowest mandatory levels to 12 to 13 m/s for both U and V components in the upper troposphere;

10 to 11 m/s for the U component and 6 to 7 m/s for V component in lower stratosphere;

and up to 20 to 22 m/s for U component and 11 to 12 m/s for V component at 10 hPa.

This means that using climatic data to check the winds, we can not detect errors smaller than 20 m/s in the best case, and errors smaller than 80 to 100 m/s in the worst case. Horizontal optimal interpolation of wind data from neighboring upper air stations produces a factor of two reduction of rms values in comparison with the climatic test. Hence, we can detect errors half as large. Figs. 6 and 7 show that vertical optimal interpolation of wind data between mandatory levels has smaller rms values than either climatic checking or horizontal optimal interpolation. The rms of vertical optimal interpolation at most mandatory levels is approximately 3 m/s (except at the highest levels, where we must extrapolated, not interpolate). This method detects errors greater than 10 to 15 m/s. The best results, as in error detection for temperature, are achieved when we use linear interpolation of wind data from significant levels. The rms values for this method are about 2 m/s. Hence, we can detect errors greater than 7 to 10 m/s. But as with temperature, every not wind observation has adjacent significant levels, and some observations at significant levels are erroneous and need to be checked.

Two other analysis procedures for winds have been developed: in the first, the real wind is compared to the geostrophic wind; and in the second, a comparison is made between the thermal wind and the vector difference of the real wind between mandatory levels. Both techniques have their own peculiarities. First, neither method can be applied in the tropics. Therefore, we can not use these methods to check the data at the many stations located between 20 degrees south and 20 degrees north latitude. Second, to apply these methods we have to calculate first and second derivatives using real observed data! Hence, these two methods are especially sensitive to the quality of the data. To correctly detect an error in a given wind observation, we must have high quality geopotential height and temperature data at neighboring stations and a reasonable distribution of these stations around the station to be checked. Averaged around the globe, except for latitudes from -20 to 20 degrees, the accuracy of the geostrophic check lies between the climatic check and horizontal interpolation of the wind (Figs. 6 and 7). The accuracy of the thermal wind procedure is slightly better than the accuracy horizontal optimal interpolation. The observed winds at adjacent mandatory levels, like geopotential height, are highly correlated, as are the observational errors. This is why better results are achieved from methods which involve a comparison of winds from adjacent mandatory levels, as opposed to methods which us horizontal interpolation. Horizontal interpolation does not take into account winds at the closest vertical levels.

The vector nature of the wind and the large values of rms for the various test for the winds make it clear that the quality control of wind data is much more difficult than quality control of geopotential height and temperature data.

It is even more difficult to detect errors in humidity data

than in wind data. The first problem is to determine which of the many possible moisture variables to test (relative humidity, mixing ratio, dew-point, etc.). Different moisture variables have different statistical distributions and are calculated using other observed thermodynamic variables (temperature, pressure) which can introduce new errors. It is not easy to decide which variable to use. One factor to consider is the availability and reliability of statistical data (means, standard deviations and correlations) for the chosen moisture variable.

We choose to use dew-point depression as the moisture variable to quality control, because statistical climatologies exist for dew-point depression and this moisture variable is usually reported in upper air observations. It is minimally affected by additional dangerous (from the QC point of view) recalculations required to convert from one moisture variable to a second variable.

The results of various checks for dew-point depression are shown in Fig. 8. The standard climatic check has an rms value ranging from 3 to 8 °C, hence we can only detect (by this check) errors in dew-point depression greater than 25-30 °C. However, this is practically the full range of dew-point depression values. Horizontal and vertical optimal interpolations decrease the rms values below 5 to 6 °C, therefore we can detect errors greater than 18 to 24 °C. However, these values are still too large to be very useful. The only method with which we can detect small errors is vertical interpolation using significant levels. Fig. 5 shows that vertical interpolation with significant levels has rms values of 1 to 1.5 °C. Hence, errors in dew-point depression data greater than 5 to 7 °C can be detected. But as before, we do not have as many significant levels as we need, and the data at these levels still need to be checked.

These poor results in checking humidity data are not very reassuring. These results are due to the large variability of humidity in the atmosphere and the micro and mesoscale nature of humidity. The upper air network was designed to observe macroscale processes. Another reason for these disappointing results is the lack of current and historical standardization of the world radiosonde network. Many different humidity sensors have been used, and the raw data from these instruments have been processed many different ways. To improve global humidity data, the hygrometers used in different countries will have to be much more accurate and compatible. Our experience in working with humidity data suggests that cloud data may be very helpful in checking humidity. But, this is a field for future research.

Finally, regarding our ability to check data at significant levels, we have a procedure that is straightforward. Data are used from mandatory levels that have been checked. This makes the logic and procedure of checking data at significant levels much easier in comparison with checking the data at mandatory levels, when we suspect both the mandatory and significant level data. A complex decision making algorithm is not needed. On the other hand, a significant level is significant because the

behavior of one or more variables changes at that level. When we check the data at a significant level, we must allow for these changes and be careful not to exclude correct observations.

Fig. 9 shows some results of interpolating temperature and dew-point depression data from mandatory levels to significant levels. The rms values for these cases are about twice greater as great as the rms values of interpolating data from significant to mandatory levels. This is due to the "significance" of the data at significant levels.

In conclusion, we can ensure a level of error detection comparable to the magnitude of observational errors for all upper air elements except humidity. This quality control system, which has been developed for the CARDS upper air data, is probably the most advanced system in existence today. There are several possible improvements to this system of complex quality control: include time interpolation of data between consecutive observations; study how to use cloud information to improve the humidity data; determine from the data a better statistical structure of the atmosphere (means, standard deviations, autocorrelations, and cross correlations in time and space).

c. Detection and removal of systematic errors

The goal of the CARDS bias program is to develop methods to identify systematic errors (which lead to biases) in the upper-air data using statistical techniques, and to develop temperature correction models for the most widely used radiosondes of the 1960-1990 period. The detailed design of a physical bias adjustment algorithm is a prominent part of this research effort.

Systematic errors can be introduced by a change of radiosonde model or data reduction method. The magnitude of the error in the upper-air temperatures has been estimated to be 1 to 3 degrees Celsius. Errors of this magnitude will be larger than any century-scale predicted climatic change for most areas. Therefore, these errors should be removed from, or at least identified in, the dataset.

Systematic errors in time series can be detected by the use of accurate station histories, mathematical-physical models, and/or statistical techniques. Unfortunately, station histories are incomplete and in many cases nonexistent. Therefore, we will have to develop statistical methods to detect systematic errors (biases) in temperature and relative humidity data. It is highly desirable to develop one or more methods that can locate a step change to within a few weeks or months of the actual date. Time series analysis will assist in determining when a station changes radiosonde models. These statistical techniques will enable us to develop rudimentary station histories for many stations which do not have written records.

Two statistical methods for detecting and correcting upper-air time-series data are being developed for the CARDS project.

The State University of New York (SUNY), under a contract with NCDC, is developing statistical methods using recursive filters

to detect step changes in time series (Zurbenko et al., 1994). This method has detected the time of documented step changes in the US humidity data to within six months of the known dates (see Figs. 10 and 11). The Alexandersson homogeneity test (Alexandersson, 1986) has been extended to detect as many as four step changes. The curves in Fig. 12 show the probability of detecting breaks using the extended Alexandersson method, the t-test, and regression. These curves have been developed from computer simulations (Rao and Porter, 1993).

A new method of detecting and adjusting inhomogeneities has been developed by Peterson and Easterling (1994) and Easterling and Peterson (1994). They used two-phase linear regression which can detect and adjust multiple step-changes in a time-difference series. This technique is based on a combination of regression analysis and nonparametric statistics. For a given candidate station, a reference series is built using observations taken at closely-related surrounding stations. A difference series is formed by subtracting each element in the candidate series from the corresponding element in the reference series. The difference series is checked for statistical changes which indicate potential discontinuities by fitting a two-phase linear regression that produces the minimum combined residual sum of squares. The significance of a discontinuity is tested using an F ratio and a t-test on the difference series. When both F ratio and the t-test are significant at the 95% confidence level, the discontinuity is saved for further testing. The difference series is divided into two parts for further checking. This procedure continues until the partitions are homogeneous or too small to test. The smallest partition tested for a discontinuity consists of 5 data points. After all the potential discontinuities are identified, the statistical significance of each discontinuity is tested using multi-response permutation procedures (Easterling and Peterson, 1994) with a windowing technique. All verified discontinuities are used to adjust the candidate series by calculating the difference in the means of the difference series before and after the discontinuity.

This method is being modified for use with upper-air data. Fig. 13 shows the results of applying the Easterling-Peterson method to monthly average data from Fairbanks, AK. The candidate series is daytime observations, while the reference series is for night observations. This method seems to correct the known bias in humidity valued for at this station during the period from 1961 to 1973 (26 April 1963 to 16 May 1972 for Fairbanks). Work is continuing in evaluating this technique.

The SUNY filter method and the application of the Easterling-Peterson method are still being studied. We must conduct further tests on these methods before either can be used to remove biases from the CARDS database.

A second type of systematic error is due to radiational, convective, and conductive effects on the radiosonde instruments and/or housing. These errors are particularly amenable to correction by mathematical-physical models. To apply physical

temperature correction models to a sounding, one must know the radiosonde type. In addition, one must know the station location, know the date and time of launch and have a good estimate of cloud height and cloud coverage.

The University of Dayton, under contract with NCDC, has developed a mathematical-physical model to correct radiosonde temperatures for radiation, conduction, and convection effects on the sensor (see Luers (1990) and Luers and Eskridge (1994)). Modelling a single daytime radiosonde ascent takes about 4 minutes on a 486/50 class personal computer. However, the model itself can be approximated by a set of lookup tables derived from several simulated ascents. The error introduced by using a fast table lookup algorithm instead of the numerical model is much less than the random observation error in the temperature readings. Lookup tables have been assembled for the VIZ, Space Data Corp., and Vaisala RS-80 radiosondes models. Work is continuing on other Vaisala models and Russian models.

Several approximation methods (Arabey (1975), Dmitrieva-Arrago and Koloskova (1969), Dolgin (1983), and Zavarina (1966)) have been used to determine cloud amounts and boundaries from radiosonde soundings. Usually the change of temperature or vertical temperature gradient and the change of humidity (relative, specific, dew-point depression or its vertical gradient) were used.

Moshnykov's method (Arabey (1975)) for predicting cloud amounts from radiosonde data from temperature and dew-point depression was improved by Arabey (1975), who developed the graphical diagram: the two-dimension dew-point depression - temperature plane is divided into four regions: first, an area of complete saturation with clouds covering 80 to 100% of the sky; second, an area not completely saturated with probable cloud amounts of 60 to 80% of the sky or with thin cloud layers; third, an area of partial saturation with probable cloud amounts of 20 to 50% of the sky; fourth, area of dry air with probable cloud amounts of 0 to 20% of the sky.

To determine cloud boundaries and cloud amounts we use the temperature, dew-point depression, and relative humidity profiles. The temperature T , dew-point depression and relative humidity U are approximated by cubic splines, and hence the second derivatives T'' and U'' are approximated by linear segments. The method assumes that clouds exist in atmospheric layers where $0 \leq T''$ and $U'' \leq 0$. In each such layer the minimum of dew-point depression is determined. This minimum value and the corresponding temperature are used to determine the cloud amount in this layer via the Arabey-Moshnykov diagram. The sky coverage for the three cloud level types (low, medium, high) are predicted as the maximum cloud amount of each cloud layer belonging to the level type.

To demonstrate that cloud boundaries can be found by the graphic and spline methods, we applied this technique to data from several radiosonde sites. Fig. 14, for example, shows the distribution of temperature, relative humidity, and their second

vertical derivatives at Brownsville, TX, 0 GMT, Jan 29, 1975. Analysis of the figure indicates a cloud layer bounded by 200 and 500 m. The Arabey-Moshnykov method predicts a single cloud layer, with coverage of 80-100%. Fig. 14 shows several other layers with $0 < T''$ and $U'' < 0$. The predicted cloud amounts are 0-20%. Cloud data from the NCDC Airways database reveal a single stratocumulus layer, with a base at 366 m and coverage of 90%. Note that a visually-measured ceiling is always higher than the condensation level, because the observer's eyes or an instrument begin to detect cloud only after the droplets' size and concentration surpass some limiting value. The difference between the observed cloud height and the condensation level can be several hundreds of meters (Shmeter, 1972).

Twice-daily radiosonde sounding data and surface-based cloud observations for 1975 to 1980 were studied at the following six stations: Brownsville, Cape Hatteras, Amarillo, Albany, Spokane, and Medford. These stations sample several different regions of the U.S. Predicted cloud heights and sky coverage were compared with surface-based observations. We selected only those cases in which the surface observer could see only one cloud layer. The results of this analysis are shown in Table III. N is the number of the observations with only one visible cloud layer, PL is the probability of correctly predicting the cloud base, PC is the probability of correctly predicting both cloud base and cloud amount for this level. The cloud amounts is predicted correctly if the observed sky coverage is within the predicted interval (8-10, 6-8, 2-5, 0-2 tenths of the sky).

Table III shows that the probability of predicting cloud level correctly is largely independent of level type (low, medium, high) and location, and on average it is more 90%. The maximum is 97.1% at Albany and the minimum is 88.1% at Amarillo.

The probability of correctly predicting both cloud level and cloud amount for each level varies with the level. At low levels it varies from 84.3% at Brownsville to 77.8% at Medford; at medium levels it varies from 89.2% at Medford to 62.9% at Albany; and at high levels it varies from 61.5% at Brownsville to 36.5% at Albany. For all levels the skill varies from 74% at Cape Hatteras and Spokane to 80% at Brownsville.

As an example of how the cloud model and the radiosonde temperature correction model will be applied, the steps needed to correct the temperature data for the VIZ radiosonde for the 1960 to 1990 period are listed below.

1. Examine station histories and apply the statistical tests to determine dates of apparent discontinuities.
2. Correct the humidity data for the 1961 to 1973 period.
3. Determine cloud heights and amounts.
4. Apply the VIZ temperature correction table lookup model.

5. Apply the statistical test to determine the effect of correcting humidity and temperature at mandatory levels.

The Equiprobability Transformation (EPT) together with empirical equations are being developed by Crutcher and Eskridge (1993) to correct the humidity measurements taken with the VIZ radiosonde during the 1961-1973 period. The empirical equations, which are fits to EPT transformed data, are a function of cloud cover and solar elevation angle.

TABLE III. Probabilities PL and PC of correctly diagnosing cloud level and correctly diagnosing both cloud level and sky coverage by the spline and Arabey-Moshnykov methods. N is the number of surface observations with one visible layer.

			CLOUD LEVELS			
			LOW	MEDIUM	HIGH	TOTAL
1	BROWNSVILLE	PL	97.5	88.4	96.9	96.1
		PC	84.3	71.3	61.5	80.0
		N	681	129	96	906
2	CAPE HATTERAS	PL	94.7	93.4	89.8	93.7
		PC	78.2	79.3	51.6	74.3
		N	495	198	128	821
3	AMARILLO	PL	85.1	88.9	100.0	88.1
		PC	78.8	70.1	61.5	74.5
		N	396	117	91	604
4	ALBANY	PL	97.0	100.0	96.2	97.1
		PC	84.2	62.9	36.5	79.7
		N	755	70	52	877
5	SPOKANE	PL	88.5	99.3	97.0	92.1
		PC	81.0	80.7	54.5	79.2
		N	331	145	33	509
6	MEDFORD	PL	94.7	96.4	98.2	95.4
		PC	77.8	89.2	61.4	79.1
		N	490	167	57	714

Huang (1993) of the University of North Carolina at Asheville has developed a computer program, based on the hydrostatic equation, which calculates station elevation from the radiosonde data. Fig. 15 shows the results of applying this program to data from Cape Hatteras, NC. Fig. 15 shows that the NCDC station histories are in error for the 1948 to 1957 period. This program can accurately determine the station elevation within 1 to 2 m.

Detection and correction of biases must wait until the

database has been compiled and the Complex Quality Control programs have been run. The following steps will then be taken to remove known inconsistencies.

1. Station elevations are known to be wrong for many stations in the database. The first systematic error that will be removed will be station elevation errors.

2. The U.S. and its cooperative stations have used a value of the gravitational constant "g", the gravitational constant different from the value used by all the other countries in the world-wide network.

3. The third systematic error to be removed is the relative humidity bias in the US VIZ data. This will be done with the techniques developed by Crutcher and Eskridge. Time series of relative humidity will be tested using the Porter and Rao methods before and after the correction to determine the effectiveness of the correction procedure.

The next step in the removal of biases in the data is more complicated. A small number of US radiosonde stations with good station histories will be selected from the CARDS database. The statistical methods that have been developed to detect sudden changes in temperature and relative humidities will be run for these stations. The statistical results, together with the station histories, will be employed to guide the use of the temperature correction model. If we succeed in removing biases from the US VIZ radiosonde data, we will test our techniques on a few well-documented stations that used Russian and Vaisala radiosondes.

If we are successful in correcting errors in stations with good station histories, we shall pick a small number of stations with poor station histories and attempt to determine and correct the detected biases.

Depending on the success of the efforts in this detection and correction endeavor, a plan will be developed for correcting the upper air data. There are a number of possibilities:

- 1) correct all data,
- 2) correct only monthly data,
- 3) correct only stations with good station histories,
- 4) correct a subset of the world-wide network,
- 5) make no corrections, but identify step changes for each station,
- 6) do nothing.

d. Database design

The CARDS database resides in a database structure created with the EMPRESS relational database management system (RDBMS). The database consists of 3 types of data: station information, metadata, and observation data. TDF63 formatted data is quality controlled via the CHQC and loaded into the CARDS database on a SUN II workstation. The CARDS station listing, load statistics, and processing tables are located on the workstation's local drives. Metadata information and observation data are stored on

a Hewlett-Packard optical jukebox, which provides 85 gigabytes of mass-media storage.

1. Station information

Two database tables are used to maintain upper air station information - CARDS station listing and station load statistics. Currently, rudimentary station history information is maintained locally and provides valid station verification during the load process.

The CARDS station listing is used to track upper air stations. One record exists for each station. Stations are identified by the CARDS number - a six digit identifier comprised of the WMO number and a sequential identifier. The station listing also includes the name of the station, the current latitude and longitude in degree/ten-thousandths, and the current elevation in meters.

The station statistics table contains an entry for each station with data currently loaded will appear in the CARDS database. Not all stations contained in the CARDS station listing may appear in the station statistics table; only those with data currently loaded. The statistics information consists of the total number of observations loaded to date, the last date information was loaded for the station, and a verification flag indicating the status of loading for that station.

2. Metadata information

One record of metadata information is maintained by observation and is stored on the optical jukebox. One table for each station is created and named according to the CARDS number (i.e. meta 720100). Metadata records are identified within each station table by the observation date and hour.

The metadata information includes latitude, longitude, elevation, observation type, termination pressure and height, availability flags for mandatory levels, various level type counts including the total number of levels, and the source of the data. This data provides an overall view of the observation.

Metadata is used to create reports pertaining to stations and blocks of stations. Inventories, source summaries, monthly summaries, and pressure summaries are all available from the metadata.

3. Observation data information

Station sorted observation data from several data sources is currently being loaded into the CARDS database. Data is stored in a separate table for each station and is located on the optical jukebox. The name of the database table reflects the station number (i.e. tbl_725650). Records are identified by the observation date and hour within each station table.

Synoptic sorted data for 1989 and 1990 has been created from

currently loaded station data. Synoptic data is maintained in tables broken down by year and month. The name of the database table reflects the year/month time period (i.e. syn_197302). This data also resides on the optical jukebox and will be the first information processed by the Complex Quality Control. Records are identified year/month/day/hour where the hour is defined as a 6 hour range. For example, hour 00 GMT for a given day includes the previous day's last 3 hours (21 GMT, 22 GMT, 23 GMT) as well as hours 00 GMT, 01 GMT, and 02 GMT.

The observation data tables are structured according to the TDF63 format. Data records consist of data header information designed to accommodate the remaining header information and the level data. The header information is comprised of the observation date and hour, latitude, longitude, and elevation, quality control effort, data source, number of levels, and the load date for each record. The text field contains 63 bytes of header, including radiosonde information and correction types applied to the data. All level information, including element quality flags, follows this header information in the text field. Each level is 56 bytes. The number of levels varies from one sounding to another.

4. CARDS database load preparation

Prior to loading the CARDS database, data files must be passed through a variety of screening processes. Generally, files are created on the UNISYS according to station number ranges and read from cartridges into the SUN.

Once read, the tape files are processed through the time duplication eliminator. This process identifies soundings which contain identical data but differ in date and/or time, within a 48-hour window.

The process discards all duplicates, retaining only the first instance of a specific sounding encountered. The process produces an output file of observations for quality control processing.

The CHQC process creates an output file of data and the corresponding metadata. Soundings with less than three mandatory levels are discarded and kept in a reject file. Records with unprintable characters or TDF63 format violations are discarded and placed in a separate debug file for further investigation.

5. Metadata load process

The fate of each sounding is determined during the metadata load process. Stations are validated against the CARDS station listing. For each new station loaded, an entry is made in the station statistics table reflecting the number of observations and the date loaded. For existing stations, the observation count and load date information is updated. Invalid station records are sent to a reject file for investigation.

The add/replace process for consolidating datasets occurs at

the metadata load level. Observations are identified by the date and hour. Because of varying practices across datasets, a two-hour window is used in identifying observations. For example, a new observation with the hour 12 GMT will be compared to observations from the same date with hours 10 GMT, 11 GMT, 12 GMT, 13 GMT, and 14 GMT. Similar logic for hour 00 GMT includes the previous day's hours 22 GMT and 23 GMT.

The logic for selecting a sounding is based upon the number of mandatory and significant levels without errors. Mandatory levels are first compared. If the number for both soundings is equal, then the significant levels are compared. If the mandatory level counts are different, the sounding with the most mandatory levels is selected. If the mandatory level counts are equal, the significant level counts are compared and the sounding with the most significant levels is selected. If, at this point, the significant level counts are the same, the currently-loaded sounding is selected. If the new sounding is selected, the old sounding is flagged for replacement and the new one will take its place in the database table.

A temporary processing table is loaded which contains observation identifiers and a processing flag. The processing flag indicates whether the sounding should be added, replaced, or discarded. Once the load process is completed and the metadata and observation data verified, the processing table is cleared.

An observation replacement log is also used to track replaced observations. Each observation flagged for replacement has an entry in the replace log. This information is used during the replace process.

6. Observation replace process

Prior to loading the observation data, all observations are flagged for replacement are extracted from their appropriate tables and are added to the replaced observation table. This table is archived periodically to keep the disk-resident portion of the table manageably small. No observation is really "discarded". Once the observation is successfully added to the replaced obs table, it is then deleted from the station database table. The replaced observation log contains a processing flag used to identify which observations have not been added to the replaced obs table.

7. Observation load process

The observation data are loaded after the completion of the metadata load and replace process. If a station has not had data loaded in the past, a new database table is created. The table is moved to the appropriate platter, indexes are created, and access privileges are granted. If the table already exists, the appropriate platter is mounted and the table is opened.

Records are compared to data in the processing table during the load process. Items checked include the CARDS number,

observation date and hour, and the data source. Each new observation is added. A duplicate of an existing observation is either added as a replacement or discarded, according to the processing flag.

8. Load verification process

Once the metadata and the data are loaded, the two sets of records are compared to insure matching metadata and data information. Upon verification, the station statistics verification flag is updated to reflect the status of the station load.

9. Preliminary results of using GTS sources concurrently

The CARDS project has been fortunate to receive data collected over the Global Telecommunications System (GTS) from three separate sources: the NMC in Washington, DC; the RRIHMI in Obninsk, Russia; and from the Australian Bureau of Meteorology. CARDS is now in the process of loading these data for the 1986-90 period. This will also be the first period processed through CQC and subsequently made available to the public.

A preliminary investigation of the benefits of using two of the three upper-air data sources concentrated on January 1990. Table IV contains the results of this investigation.

TABLE IV. Preliminary comparisons of stations and observations for NMC GTS and USSR GTS for January 1990.

	<u>Africa</u>	<u>SE Pacific</u>
Number of stations common to both sources	91	21
Number of observations - NMC GTS	2541	308
- USSR GTS	3915	579
Percent increase using USSR GTS instead of using NMC GTS	54	88
Total number of stations - NMC GTS	161	30
Number of additional stations - USSR GTS	16	15
Percent increase using both sources	10	50
Total number of observations - NMC GTS	3109	472
Number of additional observations - USSR GTS	548	318
Percent increase using both sources	18	67

It is quite evident that combining sources give a much more complete database. By adding the USSR GTS data to the NMC GTS data, we get an 88% increase in SE Pacific observations for stations common to both datasets. The USSR GTS also adds 50% more stations and a 67% increase in total observations. Each

source has its strengths and weaknesses depending on available GTS circuits, data cutoff times, etc. We look forward to seeing the final results especially with the combination of all three sources.

10. Reporting

Reporting software has been developed to help analyze the database. Current reports include metadata inventories, pressure summaries, station source statistics, and latitude/longitude band distribution. Observation data can be extracted by station for a given time range.

The metadata inventory provides the user with up-to-date information on the amount of data available for a given station. The report displays annual summaries of observation counts by month, the total number of days observations were taken, and the average termination pressure. The metadata inventory provides the user with an overall view of the data available for a station.

The pressure summary runs for a block of stations for a given year and month. The report provides a summary of the availability of data for specific levels. Records are listed by station and reflect the percentage of observations reaching each level listed. The average termination pressure for each station by year and month is included along with the total number of observations taken in that time period.

Data source statistics are available for all stations in a WMO block, or for all years for one station. The block summary lists each station for which data has been loaded and the total number of observations for each data source identified. Totals for each station and each data source are included as well. The station summary runs for one station identified by CARDS number and lists the total number of observations per data source per year, along with overall totals.

The latitude longitude distribution report provides a breakdown of the number of stations and total number of observations for a lat/lon band. Thirty-six bands have been defined.

Hour summaries are also available for a block of stations or one station. These reports provide a count of observations per hour by station. The block summary lists all stations in the block along with the total number of observations over a given year period. The station summary lists the total number of observations per hour on a yearly basis. Both summaries include overall totals.

11. Observation data retrieval

Observation data can be retrieved for any single station, for a given date range. The user must provide the CARDS number and beginning and ending dates. The data is extracted in the TDF63 format with a carriage return/linefeed following each header and

level record.

Once the data is extracted from the database, the user has four options for transferring the data to a working environment. The information can be moved to another workstation location; files can be transported across the network, via the file transfer protocol (FTP) utility, to either an off-site workstation or a PC; the data can be written to exabyte tape; or the data may be written to cartridge tape.

The user may also extract a single observation for printing and/or viewing. The observation is formatted in pseudo-TDF63 format: each level is followed by a carriage return/linefeed, and the header information is displayed on two lines.

Synoptic sorted data is extracted in day/hour files for processing in the CQC. Files are created for hours 00 GMT, 06 GMT, 12 GMT, and 18 GMT. Currently, only 1989 and 1990 are available.

e. Station histories

The first version (1.0) of the CARDS station history database was developed from three separate sources. The primary source was a database developed by Barry Schwartz, NOAA/ERL and the National Climatic Data Center. This file primarily contained information on U.S. stations, some Canadian sites, and a few, mostly military, foreign locations. The time period covered was 1945 - present. Added to this was the Master Station Catalogue compiled by the Air Force's ETAC/OL-A, co-located with NCDC. This Master Station Catalogue was a compilation of biweekly operational catalogs dating back to 1977. A December 1991 version of NMC's Master Station List was used to complete the CARDS database. The first station history contained entries for 3991 stations. This version was released in July 1993.

For the first time a comprehensive station history listing for global upper air sites was completed. Information contained for each station included WMO number (expanded to six digits for use in CARDS), station name, country or state, latitude, longitude, elevation and period of time, covered by this information. Any change in any element resulted in a new entry for that station. Table V contains an example of the station history. Further station history information can be added to this bare-bones database, i.e., instrumentation history, receiving and processing equipment, corrections applied, etc.

We soon realized that the first version contained entries for stations that were not real upper air sites. Many entries had been generated because of minor spelling change in the station name. We decided to purge the database of these extraneous stations and entries. A manual review of all 3991 stations was undertaken, comparing station numbers against Air Force Global Weather Center audits for the years 1970-90. Concurrent to this analysis of the station histories, a station history database was received from Russia for the former USSR. This information,

covering over 200 stations from the FSU for the full period of record, was incorporated into the CARDS structure. The resulting database, known as Version 2.0, was released in December 1993. The total number of stations declined to approximately 2500.

The stations that were removed from the original version of the database were compared to the CARDS observational database. It was discovered that a few hundred stations would have to be reentered.

A comprehensive digital station history for China has been received from the Peoples Republic of China. This new information covering the period 1951 to 1993 will be added to the database. The addition of the Chinese station histories and additional verification of the stations in the database will lead to a Version 2.1, which is scheduled for release in March 1994.

Table V. Data sample from the station history database for CARDS station number 725280. In this table St is the state and s is the source of the information.

CARDS number	station name	St	lat	long	elev	s	open	close
725280	Buffalo/Arpt	NY	42.93300	-78.73300	220.0	B	193909	195403
725280	Niagara Falls	NY	43.11700	-78.91700	182.0	B	195403	196008
725280	Buffalo/Arpt	NY	42.93300	-78.73300	218.0	B	196008	999912

f. Planned products

The CARDS project expects to release a worldwide dataset for the 1986 to 1990 period in the fall of 1994. This dataset will be quality controlled by the Complex Quality Control system developed for the CARDS project. This dataset will be followed by additional subsets until the entire period, 1950-1990, is released.

IV. Acknowledgements

CARDS is a joint project of the United States' National Climatic Data Center (NCDC) and the Russian Research Institute of Hydrometeorological Information (RRIHMI). The work of Oleg Alduchov was supported by the Resident Research Associateship of the National Research Council of the National Academy of Sciences and funded by NCDC. The CARDS program is supported by the U.S. Department of Energy under contract No. DE-AI05-90ER61011. The CARDS program is also supported by the Climate and Global Change program of NOAA and the National Climatic Data Center. The CARDS project is grateful to Roy Jenne and his staff at NCAR for furnishing the NCAR and MIT dataset and for help in decoding both datasets. We also thank Argentina, Australia, Brazil, Canada, China, Great Britain, Hong Kong, Hungary, Korea, Netherlands, and Russia for furnishing data to the CARDS project. We also appreciate the guidance of the CARDS Science Panel, Dr Roland Madden, chairman.

V. References

- Alduchov O.A., 1983: Complex quality control of FGGE Level II upper-air data. *Meteorol. Gidrol.*, 12, 94-102, (in Russian).
- , 1982: Combined quality control of height and temperature for isobaric surfaces in FGGE upper-air reports. GARP International Conf. on the Scientific Results of the Monsoon Experiments. Extended Abstract and Panel Session. Denpasar, Bali, Global Atmospheric Research Program, WMO, 8.15-8.18.
- Alduchov O.A., 1985: On the structure of geopotential and temperature observational errors in radiosoundings of the atmosphere. *Proceedings of AURIHMI-WDC*, issue 131, 29-39 (in Russian).
- Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.*, 6, 661-675.
- Angell, J.K., and J. Korshover, 1975: Estimate of the global change in tropospheric temperature between 1958 and 1973. *Mon. Wea. Rev.*, 103, 1007-1012.
- Angell, J.K., and J. Korshover, 1978a: Global temperature variations, surface-100mb: An update into 1977. *Mon. Wea. Rev.*, 106, 755-770.
- Angell, J.K., and J. Korshover, 1978b: Estimate of global temperature variations in the 100-30 mb layer between 1958 and 1977. *Mon. Wea. Rev.*, 106, 1422-1432.
- Angell, J.K., and J. Korshover, 1983: Global temperature variations in the troposphere and stratosphere, 1958-1982. *Mon. Wea. Rev.*, 111, 901-921.
- Angell, J.K., and J. Korshover, 1984: Comparison of tropospheric temperatures following Agung and El Chichón volcanic eruptions. *Mon. Wea. Rev.*, 112, 1457-1463.
- Angell, J.K., 1986: Annual and seasonal global temperature changes in the troposphere and low stratosphere, 1960-85. *Mon. Wea. Rev.*, 114, 1922-1930.
- Angell, J.K., 1988: Variations and trends in tropospheric and stratospheric global temperatures, 1958-87. *J. Climate*, 1, 1296-1313.
- Antsipovich, V.A., 1980: Complex quality control of height and temperature at mandatory isobaric surfaces. *Proceedings of USSR Hydrometeorological Center*, No. 217, (in Russian).

- Arabey E.N., 1975: Radiosounding data as means for cloud layers revealing. *Meteorol. Gidrol.*, 6, 32-37, (in Russian).
- Automation Division Staff, 1987: Office Note 29: NMC format for observational data (ADP reports), National Meteorological Center, Washington D.C., 49pp.
- Böttger, H., A. Radford, and D. Söderman, 1987: ECMWF monitoring tools and their application to North American radiosonde data. European Centre for Medium-Range Weather Forecasts, Operations Dept. Tech. Mem. 133.
- Buck A.L., 1981: New Equations for Computing Vapor Pressure and Enhancement Factor, *J. Appl. Meteorol.*, 20, 1527-1532.
- Collins, W.G. and L.S. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Wea. Rev.*, 118, 2752-2767.
- Crutcher H.L. and R.E. Eskridge, 1994: Development of a method or methods to modify daylight humidity data from the U.S. Weather Bureau and U.S. Air Force VIZ radiosondes for the 1961-1973 period. NOAA contract No. NE-EF2000-2-002600, 30pp.
- Dmitrieva-Arrago A.R., Koloskova L.F., 1969: Approximate method for determine cloud boundaries. *Meteorol. Gidrol.*, 6, 47-52, (in Russian).
- Dolgin M.I., 1983: Determine scheme clouds from atmosphere sounding in Antarctic Continent. *Meteorol. Gidrol.* 11, 47-51, (in Russian).
- Easterling D.R. and T.C. Peterson, 1994: A new method for detecting and adjusting for undocumented discontinuities in climatological time series. submitted to *Int. J. Climatol.*
- Elliott, W.P., 1989: Change in precipitable water. Presented at DOE Workshop: A Critical Appraisal of Simulations and Observations, May 1988, Amherst, MA.
- Elliott, W.P., and D.J. Gaffen, 1991: On the Utility of radiosonde humidity archives for climate studies. *Bull. Amer. Meteor. Soc.*, 72, 1507-1520.
- Gaffen, D.J., 1992: Historical changes in radiosonde instruments and practices, WMO instruments and observing methods report No. 50 WMO/TD No. 541, 89 pp.

- Gandin, L.S., 1963: Objective analysis of meteorological fields. Hydrometeoisdat, 287 pp., (in Russian).
- _____, 1969: On automatic quality control of current meteorological information. Meteor. Gidrol., 3, 3-13, (in Russian).
- _____, and Kagan, R.L., 1976: Statistical methods of meteorological data interpretation. Hydrometeoisdat, 360pp., (in Russian).
- _____, 1988: Complex Quality Control of Meteorological Observations. Mon. Wea. Rev. 116, 1137-1156.
- Hawson, C.L., 1970: Performance requirements of aerological instruments. WMO Technical Note No. 112, WMO No. 267, 49pp.
- Hooper, A.H., 1975: Upper-air sounding studies. Vol. 1, Studies on radiosonde performance. WMO Technical Note No. 140, WMO No.394, 1-110.
- Huang, H.J., 1992: Report on a quality control system for CARDS. Contract Number 40EANE001507, 23pp.
- Huang, H.J., 1993: Report on detecting possible errors of surface height in upper air data set. Contract number 50EANE200023 Task #1, 77pp.
- Ivanov, A., A. Kats, S. Kurnosenko, J. Nash, N. Zaitseva, 1991: WMO international radiosonde comparison-phase III, Dzhambul, USSR 1989, WMO instruments and observing methods report No. 40 WMO/TD No. 451, 135pp.
- Julian, P., 1989: NMC Personal Communication.
- Karoly, D.J., 1989: Northern Hemisphere temperature trends: A possible greenhouse gas effect? Geophy. Res. Lett., 16, 465-468.
- Kahl J.D.W., M.C. Serreze, R.E. Stone, S. Shiotani, M. Kinsey, and R.C. Schnell, "Tropospheric Temperature Trends in the Arctic, 1958-1986. JGR, 98, 12825-12838.
- Knox, J.L., K. Higuchi, A. Shabbar, and N.E. Sargent, 1988: Secular variation of Northern Hemisphere 50 hPa geopotential height. J. Climate, 1, 500-511.
- Luers, J.K., 1990: Estimating the temperature error of the radiosonde rod thermistor under different environments. J. Atmos. and Ocean Tech., 7, 882-895.

- Luers, J.K. and R.E. Eskridge, 1994: Temperature corrections for the VIZ and Vaisala radiosondes, submitted to the *J. Appl. Meteorol.*
- Middleton, W.E.K., 1969: *Invention of the Meteorological Instruments*, The Johns Hopkins Press, Baltimore, 362pp.
- Nash, J., 1984: Compatibility of radiosonde measurements in the upper troposphere and lower stratosphere for the period 1/11/1982 to 31/10/1982. *Meteorological Office O.S.M.* 24.
- Nash, J. and F. Schmidlin, 1987: WMO international radiosonde comparison final report, WMO instruments and observing methods report No. 30 WMO/TD No. 195, 103 pp.
- Neilon, J.R., 1964: Office note no. 20: Upper air ADP record format. National Meteorological Center, Washington D.C., 10pp.
- Oort, A.H. and H. Liu, 1992: Upper air temperature trends over the global, 1958-1989. To appear in the *J. Climate*.
- Parfiniewicz, J., 1976: Complex quality control of upper-air information. *Methodical Guidelines*. USSR Hydrometeorological Center, 66 pp., (in Russian).
- Parker, D.E., 1980: Climatic change or analysts' artifice?—a study of gridpoint upper-air data. *Meteorol. Mag.*, 109, 129-152.
- Peterson, T.C. and D.R. Easterling, 1994: Creation of homogeneous composite climatological reference series. Submitted to *Inter. J. Climatology*.
- Ralston, A. and C. L. Meek, 1976: *Encyclopedia of Computer Science*, Petrocelli and Charter, New York.
- Rao S.T. and P.S. Porter, 1993: Statistical methods for detecting discontinuities in time series of upper air data, annual progress report, July 1993, NOAA contract 50EANE2000078.
- Rosen, R., 1993 Personnel communication.
- Shmeter S.M., 1972: Convection cloud physics. *Gidrometizdat*, 227, (in Russian).
- Trenberth, K.E., and J.G. Olson, 1988: An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Centre for Medium-Range Weather Forecasts. *Bull. Amer. Meteor. Soc.*, 69, 1047-1057.

- Trenberth K.E., 1989: Representativeness of a 63-station network for detecting climate change. DOE Workshop: A critical appraisal of simulations and observations. Amherst, MA.
- Trenberth, K.E., and J.G. Olson, 1991: Representativeness of a 63-station network for depicting climate changes. In Greenhouse-Gas-induced climate change: A critical appraisal of simulations and observations. Elsevier, 615 pp.
- Turnor, H., 1865: Astra Castra: Experiments and adventures in the atmosphere, Chapman and Hall, London, 530pp.
- Woodruff, S.D., R.J. Slutz, R.L. Jenne, and P.M. Steuer, 1987: A comprehensive ocean-atmosphere data set. Bull. Amer. Meteor. Soc., 68, 1239-1250.
- Zaitseva, N.A., 1990: Aerology, Gidrometeoizdat, Leningrad, 325pp.
- Zavarina M.V., 1966: Determine Stratocumulus and Stratus cloud top from radiosonde observations data. Transactions GGO, No. 200, 111-118, (in Russian).
- Zurbenko, I., P.S. Porter, J. Chen, S.T. Rao, J.Y. Ku, R. Gui, R.E. Eskridge, T. Karl, 1994: Detecting discontinuities in time series of upper air data: demonstration of an adaptive filter technique. Submitted to the J. Climatol.

APPENDIX: Some dataset decoding problems

A1. The problem

As we stated in the Introduction, upper-air data have been gathered in many ways by many organizations for many decades. Each gathering system developed its own data format. As digital data processing was introduced, old data formats were digitally encoded. Several standard transmission codes existed. Different organizations encoded them differently. Storage codes did not necessarily match transmission codes. In the early days of digital data transmission and storage, data densities were orders of magnitude lower than the data densities available today. Data storage techniques were designed to save space. Data were often stored in packed binary formats. The IBM-standard 8-bit byte was not -- and still is not -- used by all equipment manufacturers. Data formats were machine-dependent. Early upper-air data managers did not know about jet streaks and similar small-scale but real meteorological features. Not all unusual values are incorrect, although some early data handling procedures may have eliminated them.

Standard-level analyses were a driving force behind data collection and archiving. Significant levels were often discarded in favor of interpolated levels at standard pressures and heights. Instrument inaccuracies may have influenced the decision to smooth and interpolate data.

As the years passed and media data densities increased, data formats changed. As instruments improved, more levels could be discerned with increasing confidence. Old data formats were abandoned, often along with their documentation. Data were not always converted to the new formats. As archives grew, the task of maintaining them grew even faster. Each archive or research organization maintaining a data store kept data in the form most convenient for it. Even so, accidents would happen. An occasional tape would grow old or be misfiled, or a manual would be lost in the cubic dekameters of the NCDC archive. A data copy procedure might appear correct, but have a fatal flaw. This year, for example, we nearly lost British Antarctic data, because of a problem with the copying of VAX BACKUP tapes on the NCDC Unisys mainframe. The Unisys tape copy job runs without generating an error condition. The copy cannot be read on the VAX. The VAX BACKUP format has not been decoded on the Unisys. This problem has occurred before, but the people receiving the data tape from Britain were unaware of it. The British Antarctic delivery tape had been logically scratched but not physically destroyed when the problem was discovered. Had we discovered the problem a few weeks later, the data would have been lost.

An employee might borrow a document and retire before it was returned. A rarely-used dataset might not receive the maintenance a primary dataset would receive. We have gaps in some of our principal datasets and inaccuracies in some of our documentation. We attempted to fill those gaps with data from our

own secondary datasets and from datasets stored by others. We encountered a variety of data formats, and a variety of documentation problems.

NCDC's principal GTS dataset, TD6103, covers the period 1973 to present. An earlier NMC-supplied global dataset ends with 1970 data. NMC GTS data from 1971-2 have been lost. Some of the worldwide data may be included in Air Force datasets, but we suspect much of the data is not. We sought additional sources. We obtained nine 6250-bpi round tapes of B-3 global upper-air data from NCAR, covering the period 1971-2 (henceforth called the ncar dataset). NCAR also supplied a 3-reel dataset originally prepared by Prof. Victor Starr at MIT, covering most of the period 1958-1963 (the MIT dataset). We cite these two packed binary datasets as examples of the decoding problems encountered in the early phases of the CARDS project.

The two datasets mentioned in this Appendix came to NCDC by way of NCAR. We wish to emphasize that NOTHING in this Appendix should be interpreted as derogatory to NCAR. The two datasets were difficult. However, had it not been for NCAR and the NCAR staff, both datasets would probably have been lost. As we note several times below, features added by NCAR saved considerable portions of the data from irretrievable loss. The added NCAR synchronization word in the NCAR-GTS ("ncar") dataset saved many soundings from the bit bucket. The NCAR-supplied documentation for the ncar dataset was superior to the official NMC documentation, ON20. Handwritten notes on the NCAR documentation and direct communication from NCAR were essential in decoding the GTS dataset. We are not attempting to assign blame to anyone. A quarter-century has passed since these datasets were compiled.

A2: The MIT dataset:

The MIT dataset is packed binary, with data at selected pressure levels (see Table A1). Some of these levels correspond to mandatory levels. Some do not. The documentation does not explain how the levels were selected: Smoothed or unsmoothed? Interpolated or selected from the original sounding? We guessed. We tagged any level at a mandatory pressure a mandatory level; the surface level, surface; and any other level, an interpolated level.

The original data from MIT were processed at NCAR, and some control information was added. Each report received a 64-bit header. Each tape block received a 64-bit block word count prefix and a 64-bit suffix, probably a checksum. The report header was well-documented. The block header and checksum were not documented. We guessed the content of the block header by inspection of hexadecimal dumps and by comparison to the ncar dataset. We do not know the format of the checksum. We used the block word count to check synchronization, and discarded the block suffix.

Table A1. Data levels for the MIT dataset:

Level#	Pressure (mb)				
1	sfc	10	600	19	150
2	1000	11	550	20	100
3	950	12	500	21	70
4	900	13	450	22	50
5	850	14	400	23	30
6	800	15	350	24	20
7	750	16	300	25	15
8	700	17	250	26	10
9	650	18	200	27	7

The MIT data conform to the documented pattern. The decoder program did not become desynchronized. Each report consists of the 64-bit NCAR-supplied header and from 1 to 27 71-bit data levels, padded to a 64-bit boundary. The header contains a report-length in 64-bit words and a level count in 71-bit levels. These counts always agreed.

The data in the levels' data fields always decoded to something: either the missing-data code or some integer. Not all bit patterns represent meteorologically reasonable quantities. The decoder converts all bit patterns to numbers, except the missing-value code. Quality control procedures will remove any non-meteorological values.

The dataset has not been quality-controlled, but a sample has passed the CHQC. The CHQC can detect errors in the MIT data, but the CHQC error-correcting algorithms may fail to correct the errors. The CHQC attempts to correct rough errors in ascii-coded decimal data: an incorrect digit, digit order inversion, etc. The CHQC decision-making algorithm was not designed to look for incorrect bits in a binary dataset. Correct binary encoding of rough errors in the original reports may be corrected by the CHQC. Errors in the binary encoding, storing, or transmission of the bit stream will not. Under some circumstances, the CHQC may actually make an unjustified correction: a change in the '8' bit might appear to be a change of one unit in the 10s place of a decimal number, for example. We hope that incorrect corrections will be infrequent.

The data were supposed to be from the period 5/58 through 4/63. Some data with dates 1957 and 1964 were encountered. The code for the year 1965 is identical to the missing-data code for the year field. Several soundings with the date 1965 were encountered, but they contained missing-data codes in all data fields. The 1965 reports were discarded. The 1957 and 1964 data were retained, subject to future quality control.

The MIT dataset used its own unique blend of units. Wind was coded in u and v components. Humidity was expressed as specific

humidity. No conversion equations to/from familiar units were supplied. Conversion of the wind was a straightforward exercise in trigonometry. Specific humidity was more challenging. Radiosonde engineering units, the raw data from the radiosonde, more closely resemble relative humidity than specific humidity. The MIT data were probably converted from relative to specific humidity by the dataset creator. We needed the inverse of the conversion algorithm. We opted to use a pair of equations from Buck (1981). If a significantly different algorithm were used in the creation of the MIT dataset, the use of Buck's equations could introduce a systematic error in the decoded data -- even if the Buck algorithm were a superior algorithm. We assume the concept of specific humidity and the physics behind its approximation have changed little in the past 30 years. We are less confident about the choice of condensed states of water used as a reference for relative humidity at temperatures below 0C. Some relative humidities computed from the MIT data differ from the relative humidities found in matching soundings in the NCDC national dataset (TD6201). We cannot explain the discrepancies.

The MIT dataset saved space by storing station location and identification information in an auxiliary table. The MIT header contains a 10-bit station number, which is an index into a station identification table prepared by the Travelers Research Center (TRC). NCDC received a copy of this table from NCAR when we got the dataset. The table was lost. NCAR supplied another copy of the table in 1992. The table was digitized at NCDC, and is now a part of the CARDS MIT dataset decoder. Some discrepancies in station history information between the TRC table and the NCDC master station list have been noted. These discrepancies will be resolved later. We suspect that the space-saving measure may have introduced inaccuracies into the data. Kahl et. al (1992) used the MIT dataset in a study of temperature trends in the Arctic, north of latitude 65N. They report: "...in terms of temperature data, one is also left with the distinct impression of a frequent mismatch between the soundings and station headings. Indeed, some of the profiles resemble those of midlatitude sites. We excluded MIT soundings in the initial computation of temperature trends." Mid-latitude soundings could be assigned to stations north of 65N if there were rough errors (naughty bits) in the station number field. The CARDS Complex Quality Control horizontal check and time-series check will be able to detect this type of error, but will not be able to correct it. To the extent that rough errors contaminated the station number field at any stage of the data processing, the space-saving scheme had an unanticipated success. Space is definitely saved when a mislocated sounding is purged from the database.

The new NCDC upper-air format, TDF63, includes station identifier, latitude, longitude, and elevation in each report's header. We wish to reduce the likelihood of future data loss caused by misidentification of stations. The geographical information and the station identifier can be used in a mutual

consistency check. The probability of simultaneous rough errors in both fields is very small.

The MIT database may contain important data from data-sparse regions. Dr Richard Rosen was a student at MIT when the dataset was being used there. He told the CARDS team that the MIT dataset contains data from Africa and other regions probably missing from the United States' principal datasets (Rosen, 1993). We cannot choose to avoid the difficulties involved with this dataset without risking the loss of data unobtainable from other sources. Efforts to use the MIT dataset will continue.

A.3: NMC global B-3 data obtained from NCAR:

The 1971-2 B-3 dataset from NCAR was difficult to decode. Documentation is substandard. Portions of the documentation are contradictory or false. Large portions of the data are not documented. Extracting clean soundings from this dataset was high as difficult as extracting clean ponies from the Augean stables -- and for very similar reasons.

The ncar dataset was apparently assembled at NCAR from data supplied by NMC. The documentation claims the NMC data were written in NMC Office Note 20 (ON20) format (Neilon, 1964), in 36-bit words. NCAR attached a 64-bit identification word to each report. NCAR blocked the data and added a 64-bit prefix word and a 64-bit suffix word to each block. Even with these added synchronization features, the decoder program became desynchronized on about 0.1% of the blocks and had to use the physical inter-block gap on the binary tape to re-synchronize. We suspect the difficulty is related to undocumented data, comprising about 80% of the data on each tape. ON20 contains a table listing many kinds of data which might be present in ON20-format datasets. The NCAR-supplied documentation and ON20 itself only describe radiosonde data. The ON20 data header contains a bit field coding for other data types, including aircraft observations. When we examined hexadecimal dumps of ncar data near the point of decoder desynchronization, we found: the previous header record usually contained the code for aircraft observations; the data were definitely not in the format documented in ON20; the NCAR-supplied data length in the header did not point to another NCAR-supplied header. While the loss of aircraft observations is not a problem for the CARDS project, the lack of a well-defined data format for aircraft and other observation types is. We cannot bypass a report without knowing its length. We cannot get its length without knowing where the length is stored and how it is encoded. When the NCAR-supplied data-length word does not point to another data-length word and there is no other measure of report length, we have no option but to discard the remainder of the tape block. The discarded data could contain valid soundings, but it cannot be decoded unless it can be re-synchronized. The binary data contain no unambiguous synchronization fields: all bit patterns are potentially legitimate data. Undoubtedly, an artificial-intelligence routine

could be devised to locate probable report headers. However, such a task would be uneconomical for processing a nine-reel dataset in a thousand-reel project. Thus, the lack of documentation of data we do not want led to the probable loss of data we do want.

Largely for this reason, we included the synchronization character '#' in the header of the new TDF63 format. Data desynchronization in TDF63 can affect more than one report only if the synchronization character is altered and the data-level count stored in the previous report is incorrect. Even if this were to happen, synchronization could readily be reestablished with the next report's synchronization character.

The data decoding problems continued even after the non-radiosonde data were stripped and a valid sounding was parsed out of the bit stream.

Each report has three header fields: the NCAR-supplied doubleword, a 108-bit NMC-supplied identification field, and a 72-bit NMC-supplied station elevation and significant-level count field. The NCAR and NMC headers appear at the start of the observation. The elevation/count field begins at bit 1793. Between these two header sections, there are 15 108-bit mandatory pressure level fields. The number of mandatory level fields does not vary, so the elevation/count field can be found easily. The elevation/count header gives the counts of various types of significant levels, in 36-bit NMC words. Note that 36 bits is 4 1/2 bytes, an inconvenience on a byte-oriented machine. Some significant levels use 1 word; some take 2 words; and others, three. There are many opportunities for inconsistent counts. The NCAR-supplied report length has proven to be reliable. However, an occasional report must be discarded due to inconsistent internal bit counts. The NCAR-supplied report length was used to re-synchronize. The elevation/count field also contains an 18-bit radiation correction indicator field. This field is undocumented in ON20 and in the NCAR documentation, and NCAR was unable to provide any help decoding it. Although we would like to have information about radiation corrections, we chose to discard the undocumented field.

The fifteen positional mandatory levels are easily parsed. Unfortunately, the heights are given in D-values, differences from standard heights. No standard height table was supplied in the original NCAR documentation. ON20 has no standard height table, either. We requested and obtained a standard height table from NCAR, but have limited assurance that it is the same height table originally used at NMC (see Table A2). An incorrect table would introduce systematic errors into the heights. A sample of the data pass a hydrostatic check when the NCAR-supplied standard heights are used. Hydrostatic consistency is a necessary but not a sufficient condition for the correctness of the standard height table. Horizontal checking of NCAR soundings against nearby TD6201 soundings should help answer the question. The CARDS Complex Quality Control procedure will perform this check. Were the table to be incorrect, the CQC might detect but would not be

able to correct the resulting height errors.

The significant temperature levels were easy to decode. The documentation was correct.

The tropopause level fields were easy to parse. The 4-bit tropopause characteristic and the 5-bit wind characteristic fields were discarded, since we lack the documentation to use them. An ambiguity in documentation caused problems with wind decoding. The standard field-missing code in the NCAR dataset is: all bits except the sign bit on.

Table A2: ncar mandatory level pressures and standard heights:

Seq#	pressure, mb:	Std Ht, Dekaft:
1	1000	37
2	850	478
3	700	988
4	500	1828
5	400	2356
6	300	3005
7	250	3398
8	200	3866
9	150	4468
10	100	5317
11	70	6065
12	50	6769
13	30	7839
14	20	8688
15	10	10188

The missing-value code for wind direction is identical to the code for 310 degrees, if wind direction is considered a signed quantity. Nothing in the documentation indicated whether wind direction or any other quantity was permitted to be unsigned. We decoded direction as an unsigned quantity, but used the documented missing-value code. Therefore, our decoder discarded all tropopause wind directions of 310 degrees. Quality control picked up the inordinate number of missing tropopause winds. We asked NCAR for clarification, and ran an empirical study. The empirical study showed the absence of wind direction 310 degrees in the decoded output. We turned off the trap for missing direction. Wind direction 310 returned. A few cases of wind direction 630 appeared, corresponding to all bits on in the direction field. Some wind speeds of about 140 m/s also appeared. NCAR sent clarification. The value 31 in the wind direction field only represented missing when the wind speed field also contained its missing-value code: 255 kt, about 140 m/sec. In addition, the missing-value code for wind direction or speed could be as documented (all but the leading bit on), or it could be all bits including the leading bit on. These extra

missing-data codes were not documented.

The winds-at-height levels were easy to decode.

The levels called "wind by pressure" (Wp) in the documentation caused considerable problems. The memo attached to ON20 describing changes to this field was incorrect. The memo states, "The coded form (3 groups of five digits each separating blank character) of the levels of 7, 5, 3, 2, and 1 mb will be carried in the section now referred to as wind by pressure (Wp). ...". The coded form of these levels was left as an exercise for the student. A hint was provided: ... if the data is complete with winds for all five levels, the (word) count will be 15. No evidence of 7, 5, 3, 2, or 1 mb levels was found in the actual data. The NCAR-supplied documentation was essentially correct, but incomplete. NCAR suggested:

"...but it seems the replacement is the coded form
(1NNPPP 1TTDD 1ddff) of tropopause level, maximum wind
level, and surface level."

The NCAR-supplied documentation did guide us in the correct direction: code books for specially-coded blocks from the standard Teletype transmission code. The Wp section contains 108-bit levels. Each level is composed of three six-character "external BCD" groups from the ADP transmission report. "External BCD" is not defined. A computer encyclopedia in the NCDC library had no reference to "external BCD", but did have the "extended BCD" format. (Ralston and Meek, 1976). External BCD was quickly shown not to be extended BCD. NCAR was able to supply a conversion from external BCD to familiar ASCII characters. Decoding could begin. We made some assumptions, again with the help of NCAR, to decode the occasional group in a form similar but not identical to the form handwritten in the margin of the NCAR-supplied document. For example, several 6-character groups had a leading "2" instead of "1". We have reasonable confidence that most of the Wp section is being decoded correctly, and hope the Complex Quality Control algorithm will expunge the parts not properly decoded.

A number of other problems arose in the decoding of the dataset. Some reports contain station identifiers in the external BCD code mentioned above. Decoding of these identifiers was deferred until we received the code table from NCAR. Tropopause levels can be found in the Tropopause section, the Temperature Significant Level section, the Wp section, and occasionally in the Mandatory Level section. Even if the data were intended to be identical, differing field widths and units can cause differences in the data. For example, the Tropopause section codes wind direction in tens of degrees. The Wp section has a field width sufficient for direction by whole degrees, but actually contains direction to the nearest five degrees. The Wp section uses a temperature sign convention: tenths digit odd means negative temperature. If this convention applies to the Significant Temperature section or the Tropopause section, it is not documented. If it doesn't, multiple values of tropopause temperature could be present. Tropopause section heights are in

hundreds of feet. Other sections which include height give height or D-value in tens of feet. The decoder has no way of determining which of these possible multiple wind, temperature, or height values is correct, so it attempts to carry them all. The result is the duplication of some levels in the decoded sounding. Similar problems have arisen with several other types of levels.

After the data had been decoded, we discovered a problem with the date on observations with time 23 GMT (Section III-1). The original date field, if any, had been stripped from the soundings. The date and hour are found in the NCAR-supplied 64-bit header. A given day's operational data file for 00 GMT will contain the 23 GMT ascents from the previous day. Radiosondes are released before nominal observation time, because a complete ascent may take 2 hours. Radiosondes released between 2231 GMT and 2330 GMT are correctly coded as 23 GMT ascents. They are used in the subsequent 00 GMT synoptic analysis. The NCAR header contains "today's" date and the time 23 GMT for "yesterday's" 23 GMT ascent. The sounding is interpreted as if it were launched 24 hours after its actual launch time. This may have been an unfortunate artifact of the original B-3 tape coding conventions. In the 1973 NMC ADPUPA format, NMC Office Note 29 (ON29, Automation Division Staff, 1987) and NCDC dataset TD6103, individual soundings do not contain dates. The date is found in a header record at the start of each tape data group. If a similar convention were used on the older data format, the substitution of the incorrect date can be easily understood. The CARDS team chose to include date, nominal hour, and release time fields in each TDF63-format sounding's header to reduce the chance of this error's occurrence in the CARDS dataset and in future NCDC datasets.

When the CARDS project has completed the translation of difficult datasets like the MIT and NCAR datasets into the standard TDF63 format, researchers will be able to access the data quickly and easily. We trust the damage we have done to the original data will be no greater than the damage likely to be done by an ad hoc decoder written quickly by a busy investigator. We expect that this damage, if any, will be detected and possibly corrected by the Complex Quality Control procedure to be run on the entire CARDS dataset. The conversion of diverse and user-unfriendly data formats into one easily-accessed format is a major accomplishment of the CARDS project. It will save investigators many man-years of research effort, while providing access to datasets previously judged unusable. Investigators face too many other challenges to be saddled with the task of dealing with balky, obsolete data formats.

Figure 1. U.S. upper-air sounding network form 1898 to 1945. This graph is from the National Climatic Data Center archive.

Figure 2. Map showing distribution of radiosonde and pilot balloon stations in 1980.

Figure 3. The number of stations taking upper air observations according to the CARDS station history, Version 2.0.

Fig. 4. RMS values of the real residuals of variable methods of predicting height, H, and thickness, h, at mandatory levels.

Fig. 5. RMS values of the real residuals of various methods predicting temperature, T, at mandatory levels.

Fig. 6. RMS values of the real residuals of various methods predicting the zonal wind, U , at mandatory levels.
(89/01/15/00, global set of 759 stations)

Fig. 7. RMS values of the real residuals of various methods predicting meridional wind, V, at mandatory levels.

Fig. 8. RMS values of the real residuals of various methods predicting dew point depression, DPD, at mandatory levels.

Fig. 9. RMS values of the real residuals for various methods predicting upper air elements (T, DPD, U, V) at significant levels using linear interpolation (LI) from mandatory of data from mandatory levels.

Figure 10. Plot of day-night specific humidities at 750 hPa for Amarillo, Texas from July 1957 to December 1990.

Figure 11. Plot of filtered data from fig. 10. The Zurbenko filter was applied 5 times using a window of 731 days. It shows a change in March 1964 and April 1973.

Power

Alex power Regr. t test

—*— —◆— —★—

Figure 12. Probability of detecting a discontinuity of size 0.6 standard deviations using Alexandersson's method, regression, and the t-test.

Figure 13. Plot of monthly average night-day humidity values during July from 1949 to 1990 for Fairbanks, Alaska. The bottom figure (b) shows the corrected data after applying the Easterling-Peterson method.

Fig.14. Distribution of temperature and relative humidity and their second derivatives. (Brownsville, 1975/01/29/00).

Figure 15. A time series of computed median surface height (solid curve), surface heights from station history (thin dashed line), and the surface height used in the radiosonde sounding for Cape Hatteras, North Carolina, WMO number 72304.