10A.4

GATHERING CLIMATIC DATA OF THE HIGHEST QUALITY

David A. Robinson
Rutgers University, New Brunswick, NJ

1. INTRODUCTION

The need to have climatic data of a high quality has received increased attention in recent years as the utility of this information in applied studies has multiplied and concern over human impacts on the climate system has grown (Robinson, 1989; Robinson, 1990; Shea et al., 1995; Karl, 1996). This is true for climatic data gathered from satellite platforms and balloons, as well as at and under water, land and ice surfaces. Careful measurement and documentation is required to acquire the best possible information and make it available for use.

For most purposes, especially in studies of climate extremes, variability and change, the data should be gathered for lengthy intervals, using accurate instrumentation and consistent observation methodologies. All aspects of the data gathering process should be documented. Such documentation has become known as metadata, or data about data. All data should be made available in digital format once quality control measures have been applied.

Gathering accurate climatic data is a difficult task. Anyone planning to collect data or use a climatic database must be cognizant of potential data inaccuracies and biases that may result from problems at any stage of the acquisition and archiving process. Simply put, you must "know" your data. To illustrate this point, this paper will outline some of the requirements for obtaining the best possible data at continental observing stations. How closely these requirements have been adhered to should be questioned before using the climate data gathered at these sites. Two examples of subtle problems related to temperature data at one observing station will illustrate the need for the meticulous scrutiny of this information before it is used. The examples will also serve to emphasize the importance of using multiple sources of data whenever possible.

2. OBSERVATIONS

Climatic data must be accurate and comparable to other, independent, observations gathered at that site at another time, or at another station. This requires accurate instrumentation, a standard site exposure, and a competent observer. Instrumentation may vary from one climate network to another or even within a network (i.e. the U.S. Cooperative Program uses liquid in glass thermometers at some sites and Maximum Minimum Temperature System (MMTS) thermometry at others). However, all instruments should meet specified accuracy standards, and intercomparisons are essential to assure that any differences between instruments are understood (i.e. Quayle et al., 1991; Blackburn, 1993).

Efforts must be taken to adhere as closely as possible to station siting regulations. This includes avoiding things such as placing a sheltered thermometer over an artificial surface or in a shaded location. A competent observer is required to make sure that instruments are functioning properly, and that observations are being gathered using correct methods and at prescribed times.

To meet these standards: 1) instrumentation must be selected carefully and evaluated continually, 2) stations and station data must be inspected frequently by professionals, and 3) observers must receive thorough and ongoing training. Any problems should be recognized quickly and immediate remedial efforts undertaken. For instance, a broken thermometer should be replaced, an observer making seemingly unrealistic observations should be spoken with, and a site that no longer meets exposure criteria should be adjusted (i.e. a tree trimmed) or, if absolutely necessary, the station should be moved.

3. METADATA

The many facets of the data gathering effort must be documented meticulously. Without such information the user of the data is likely to misuse or misinterpret the information. Quality metadata includes information on such things as: 1) site location, 2) surface conditions at and surrounding the observing site, 3) instrumentation, 4) time of observation, and 5) name of the observer. This information has been gathered for stations within many observing networks, although in some cases not as thoroughly as desired. Recent efforts have begun to make metadata, past and present, more readily available to the data user. Innovative ways of gathering information about site locations are also being developed (K. Gallo, per. comm.). Managers of metadata information files should also consider providing information on known errors or inconsistencies related specifically to the station, including problems that have been recognized over the

*Corresponding author address: David A. Robinson, NJ State Climatologist, Dept. of Geography, Rutgers University, New Brunswick, NJ 08903; email: drobins@rci.rutgers.edu.

entire time the station has been in existence. This would complement more general network information regarding differences between instruments, factors needed to adjust for changes in observation time, etc.

4. DATA ARCHIVES

Climatic observations and metadata of the highest quality must be collected, scrutinized, and made readily available at professional archives. These archives may vary in size and scope, ranging from data for a specific experimental program on-line at an investigator's institution, to a government archive, such as the U.S. National Climatic Data Center (NCDC). For recent observations, most datasets are available in digital format, although extensive keying of historic data remains to be accomplished. The first key to building a quality database is to minimize errors made in transferring data from the observing site or from individuals gathering metadata to the archive. This is increasingly being accomplished by electronic means, however it continues to include such routines as double keying data from Cooperative Observing stations received at NCDC.

Once data are at the archive, extensive quality control of the observations must take place (Reek et al, 1992; Peterson, 1993). This may involve internal comparisons of variables gathered at a specific station on a given day, to adjacent days, or to a station climatology. It may also include a comparison of an observation to others gathered simultaneously at nearby, comparative locations. Quality control at surface observing stations is most often done using daily data. In the next section, subtle problems with observations that may not be apparent when making daily checks are discussed.

Finally, data and metadata files must be available to potential users. In recent years, the means of making data available have changed considerably, and will likely continue to evolve rapidly. Paper copies of data have moved to punch cards, then to nine-track tapes, and on to today's higher density storage media. The past few years have seen the introduction of electronic transfer of datafiles, using file transfer protocol (FTP) and the World Wide Web.

5. IDENTIFYING SUBTLE OBSERVATION ERRORS

As mentioned earlier, subtle changes or persistent problems at a station may occur over months or years, and not be identified through daily quality control efforts. Quite often the best method to identify such errors is to intercompare monthly or annual totals, means or other statistical variables from nearby, climatically similar sites. At least three stations must be compared in order to identify a station with a problem.

Recently, an evaluation of the quality of past and present data being gathered at stations in New Jersey identified two types of temperature errors at Long Valley, a National Weather Service (NWS) Cooperative Observing station located in the west central part of the State. In the first case, mean monthly minimum temperatures were compared with those at the nearby (no more than 32 km distant) Belvidere Bridge and Flemington, NJ Cooperative stations. Monthly differences in minimums between Long Valley and each of the other stations were calculated, and a ten-year portion of the results is shown in figure 1. Of immediate note is the sawtooth shape of the time series, indicating an unstable relationship between minimums at Long Valley and the other two stations. Since both other stations have virtually the same relationship with Long Valley, and there is no natural climatic explanation for the variations, there must have been a problem with the thermometer at Long Valley during this period.

While it has been impossible to fully verify the problem at the station, the situation appears to be as follows. The minimum thermometer at Long Valley is liquid in glass, an instrument which is known to occasionally exhibit liquid separation within the measurement column. Apparently this separation began to occur gradually at Long Valley during the mid to late 1980s. Before being recognized midway through April 1988 (month 40 in figure 1), minimum temperatures at the station were reported as being close to 6°F colder than the other sites, whereas in 1985 they had only been about 2°F lower. It appears that rather than replace the thermometer, it was "shaken down" so that the column again merged and the mean minimum at Long Valley was once again only a few degrees lower than the other sites. However, immediately after this was done, the column began to separate once again. Only this time, rather than taking about three years to separate four degrees, it took about four years to separate three degrees (it is possible that this was a new thermometer, that happened to have the same problem as the first). Again, this problem was eventually recognized, and it is known that the thermometer was replaced during November 1992. Since that time, Long Valley minimums have most often been from 0 to 2°F cooler than the other sites.

It is unfortunate that it took so many years to recognize the problem each time the thermometer began to separate. Some of this is understandable, as the drift in temperature was subtle and slow. However, eventually the offset was large enough that it should have been recognized in periodic evaluations of daily data by the observer (rain falling and not freezing on surfaces while the thermometer read several degrees below freezing), by NWS personnel responsible for the maintenance of the station (if required annual inspections of the station occurred), and by quality control routines at NCDC (the station reported the lowest daily NJ minimum for the month five times in the first 10 months of 1992, and was one degree off the

State minimum four times; this is climatically uncharacteristic for the site).

A second temperature problem noted at Long Valley can not be easily recognized by evaluation of daily temperatures at NCDC, but should not be permitted to exist according to siting specifications. As with the minimums, the mean monthly maximums at Long Valley were compared with Belvidere Bridge and Flemington maximums between 1985 and 1994. A seasonal fluctuation in the differences between Long Valley and the other stations is noted throughout the ten-year study interval shown in figure 2. A visit to the stations found that the Long Valley instrument shelter sits in the shade of a deciduous tree, while shelters at the other sites are exposed to considerably more sun during the warm portion of the year. Thus maximum summer temperatures are considerably cooler at Long Valley, while during the leafless winter they lie quite close to maximums at the other stations, as expected from their nearby locations. The shaded exposure does not meet siting standards, making direct comparisons of the Long Valley summer maximum records with other official observing stations impossible. NWS inspectors should not have permitted this siting, or should have moved the shelter once shading became a problem. This, admittedly, is a difficult situation, as observers are not always easy to recruit and retain, and observers may not wish to have a shelter placed in a conspicuous sunny location on their property.

6. CONCLUSIONS

Accurate climatic data is an extremely valuable resource. Considerable efforts are necessary and justified to assure that information of the highest quality is collected and made readily available to the ever-growing user community. This begins by having accurate and properly situated instruments, well-educated observers, and diligent management of observation locations. This must be followed by careful scrutiny of the data gathered, and documentation of station particulars, including known problems with the data. Finally, the data and metadata are ready for dissemination.

Examples of difficulties in meeting observational goals have been presented. The intent was not to malign a particular station or individuals associated with its operation; all stations and observing networks have problems from time to time. A data user must look carefully at all information pertaining to a station before using it in a study. Better yet, whenever possible, a user should employ data from multiple sites or sources. Fortunately, this is an option that is often available in this era of multiple surface and satellite networks.

Acknowledgments: Thanks to the New Jersey Agricultural Experiment Station, Rutgers University, for supporting the Office of the NJ State Climatologist, and to N. Stefano, Sussex 8NNW Cooperative Observer, for assistance with the Long Valley evaluation.

REFERENCES

Blackburn, T.M., 1993: Effects on climate change resulting from changes in National Weather Service cooperative station instrumentation and siting conditions. *8th Conf. Appl. Meteor.*, Anaheim, CA, AMS, J6-11.

Karl, T.R. (ed.), 1995: *Long-Term Climate Monitoring by the Global Climate Observing System.* Kluwer Academic Publishers, Dordrecht, 518pp.

Peterson, T.C., 1993: Report of the International Workshop on Quality Control of Monthly Climate Data. *NCDC Global Climate Laboratory Monograph, ORNL/CDIAC-69*, 8pp & appendices.

Quayle, R.G., D.R. Easterling, T.R. Karl, and P.Y. Hughes, 1991: Effects of recent thermometer changes in the cooperative station network. *Bull. Am. Meteor. Soc.*, **72**, 1718-1723.

Reek, T., S.R. Doty and T.W. Owen, 1992: A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network. *Bull. Am. Meteor. Soc.*, **73**, 753-762.

Robinson, D.A., 1990: The United States cooperative climate-observing systems: reflections and recommendations. *Bull. Am. Meteor. Soc.*, **71**, 826-831.

Robinson, P.J., 1989: Climate Information Needs in the Southeast United States. *Final report: NOAA Grant No. NA88AA-D-CP041 to the Southeast Regional Climate Center*, 36pp.

Shea, D., R. Jenne, and W. Spangler, 1995: Daily and monthly maximum and minimum temperature datasets archived at NCAR and some applications. *Atmospheric Research*, **37**, 11-17.

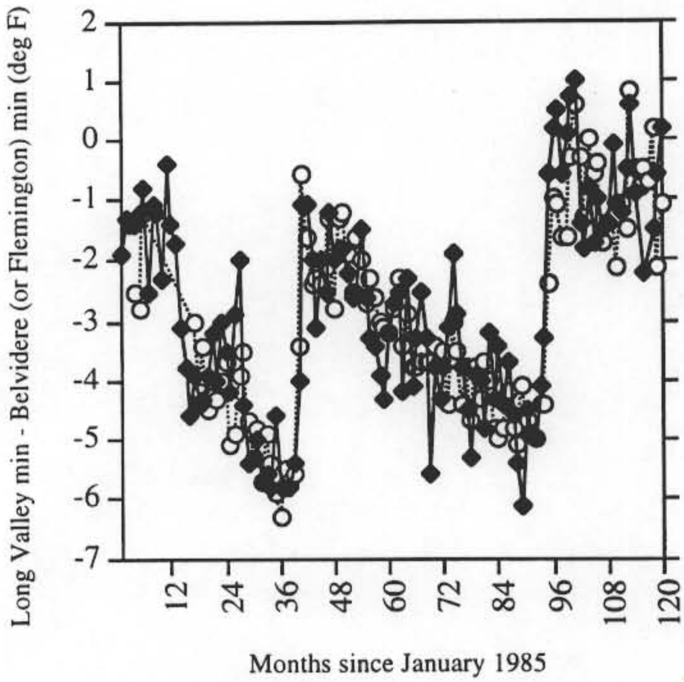

Figure 1. Differences in monthly mean minimum temperatures between Long Valley and Belvidere Bridge (diamond) and between Long Valley and Flemington (circle) from January 1985 to December 1994. Negative numbers indicate Long Valley temperatures are colder than the other stations.

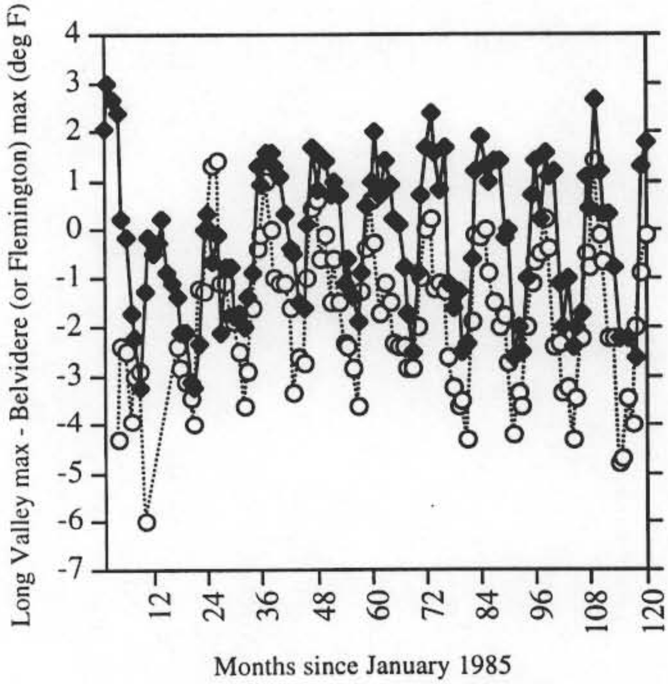

Figure 2. Same as figure 1, except for monthly mean maximum temperatures.