J3.20 A NEW PARADIGM IN NEAR REAL-TIME COOPERATIVE DATA INGEST AT NOAA'S NATIONAL CLIMATIC DATA CENTER

Karsten Shein * and Timothy W. Owen
NOAA National Climatic Data Center, Asheville, North Carolina

ABSTRACT

As the Nation's official archive for climate information, the National Oceanic and Atmospheric Administration's (NOAA's) National Climatic Data Center (NCDC) is charged with preservation and accessibility of a quality digital record of Cooperative (COOP) Network data and metadata. This record has historically been derived through the imaging and keying of so-called 'B-91' forms that are sent by observers, via the National Weather Service (NWS), to NCDC. The processing time, including quality assurance checks and serial publication, typically is 45-60 days from the end of the data month.

Technological and communication advances, coupled with integrated climate, weather and water reporting needs have reached a point where near real-time (*i.e.*, daily) reporting of observations is feasible. While ASOS data have long been directly reported to NCDC in this time horizon, COOP data reporting is delayed by the continued use of monthly paper forms. Rapid data reporting is fundamental to the success of the U.S. effort in Global Earth Observations, especially for such time-sensitive activities as the drought monitoring of the National Integrated Drought Information System (NIDIS). Additionally, improving reporting frequency is important to the transition of Legacy COOP under NOAA's Environmental Real-Time Observing Network (NERON, 2006).

NCDC is coordinating with its Regional Climate Centers (RCC) and the NWS to develop a Web interface based on existing systems (*e.g.*, Weather Transmitted Cooperative Observer Data Encoded Report (WxCoder), Community Collaborative Rain, Hail and Snow Network (CoCoRAHS) and Environment Canada's Cooperative Online Temperature and Precipitation Entry System (COOLTAP)) to facilitate the electronic submission of COOP data to NCDC, and thus to the climate community on a daily basis. To this end, the following guiding principles have been identified:

1) Provide an efficient, *easy-to-use data entry system* for participating COOP observers,
2) Ensure *timely* availability of COOP data to all customers,
3) Improve data quality through *automated near-real-time data quality control (QC)*,
4) Achieve a *near-paperless* electronic data collection, transmission, and archiving system.

* *Corresponding author address:* Dr. Karsten Shein, NOAA National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801; e-mail: Karsten.Shein@noaa.gov.

1. INTRODUCTION

NOAA's NCDC is designated through the Federal Records Act of 1950 as the archive for the long-term preservation of the Nation's climate record. NCDC also has responsibility to provide "useful and readily available climate information on a continuing basis" (National Climate Program Act, 1978). Taken together, NCDC continually endeavors to provide efficient and convenient access to the highest quality data possible in a timely fashion.

Of its broad data holdings, the *in-situ* daily observations of temperature, precipitation and snow that are reported by the National Weather Service's (NWS) Cooperative Observer (COOP) Network are among the most requested. NCDC dedicates significant resources to the thorough quality assessment and control (QA/QC) of these data, including the Automated Surface Observing System (ASOS) locations, and presents these data in several popular serial publications, such as *Local Climatological Data (LCD)*.

Technological solutions have been leveraged to improve the QA/QC processes and in transitioning to the digital production of serial publications. However, the effectiveness of such improvements is hindered by the continued practice of monthly, paper-based climate observation reporting. Despite the availability of several NWS tools (described below) for more timely data reporting from COOP observers, the network continues to be reliant upon a majority of observers who provide the NWS with handwritten observation forms (*i.e.*, B-91 forms). These B-91 forms are submitted to NWS field offices only at the end of the observation month, manually checked by NWS personnel, and forwarded to NCDC, where they must be digitized by data entry contractors and subjected to QA/QC processing before they become accessible to the climate community. The resultant lag in final data release is between 45-60 days from the end of the data month, which is clearly too late to support a host of climate monitoring needs.

2. DATA ENTRY SYSTEMS

Both telephone and internet-based systems have been developed for daily weather data entry by COOP observers. The NWS Remote Observation System Automation (ROSA) telephone entry system allowed for several hundred stations to report their observations on a near real-time (*i.e.*, daily) basis in the 1990's (NRC, 1998). Technological advances in telephony led to the replacement of this system with the Interactive Voice-Remote Observation Collection System (IV-ROCS) in 2003. Both of these systems were deployed primarily on a regional basis, and did not supercede the paper form submission process. Likewise, an Internet-based data entry system called WeatherCoder (or WxCoder) was

developed (NWS, 2003). The advantages of a web-based approach included visual verification, simplified coding and a significantly reduced chance of digit transposition. WxCoder allowed for basic internal consistency QA/QC as well.

The Meteorological Services of Canada were successful in deploying a national web-based data entry system (called COOLTAP) in 2004. In the first year, Environment Canada was successful in migrating the official observations of nearly a third of all observers into this new paperless system (with a telephone backup).

Currently, a rapidly expanding volunteer network, the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS), is wholly dedicated to web-based data entry. Formed in 1998, this Colorado State University activity has expanded to 2,500 observers in 12 states (Colorado Climate Center, 2006).

3. DEVELOPMENT OF WxCODER III

Building upon the experiences of previous data entry systems, NWS and NCDC, in cooperation with the Regional Climate Centers (RCC), are developing a nationally-supported Web-based data entry system called WxCoder III. Unlike previous efforts, this system will have national projection and will supersede B-91 paper form submissions.

The Western RCC in Reno, Nevada is playing a leading role in the development of this new interface and data management system that will leverage upon the RCC Applied Climate Information System (ACIS). ACIS is a distributed data infrastructure that ties data collection sources with end users, while being synchronized to the NCDC database.

An early mock-up of the WxCoder III interface is provided in Figure 1. In keeping with the principles articulated in the Abstract, the interface is presented in an easy-to-use format that can be customized to fit the needs of the observer. The interface will incorporate many of the successful features of COOLTAP, including features to allow the observer to graphically view past data.

4. NCDC PLANS FOR PAPERLESS DATA INGEST

In addition to the decreased interval between observation and receipt by NCDC, a fully-realized, Web-based COOP observer submission process represents a significant improvement over the existing COOP data ingest at NCDC. Of the aforementioned NCDC components of the COOP data capture methodology, the digitization of manually-recorded observations and their subsequent quality assurance processing consume a large amount of human and computing resources.

Eliminating the need for manual digitization of paper forms serves two purposes for NCDC. First, it drastically reduces the time between when an observation is recorded and when it is made available to the climate community. With paper B-91 observation forms, the interval between observation and availability may be as much as three months. The use of a Web interface by an observer means that a digital version of a day's observations can be ingested by NCDC within seconds of submission. Secondly, although NCDC protocols minimize the possibility, the process of digitizing manually-recorded observations may introduce key-punch errors into the digitized data (*e.g.*, misrecording of illegible entries, *etc*.). A completely digital process, from observer to access, removes the potential for such key-punch errors.

Additionally, because observations are submitted by the observer in a digital format, certain automated QA/QC algorithms may be applied to the observations at the time of submission. Such QA/QC, known as internal consistency checks, check that the values being entered into the Web form are consistent with the

Figure 1. Mock-Up of Main Web Entry Page for WxCoder III (Western Regional Climate Center)

expected values of such observations. Common reasons that data may fail internal consistency include transposed digits, erroneous decimal placement, or values entered into the wrong meteorological element (*e.g.*, maximum and minimum temperatures are reversed). With such QA/QC built into the Web interface, most internal inconsistencies can be eliminated before the observer is even permitted to submit their observations.

Another set of issues faced by NCDC QA/QC is temporal. First is the relatively common problem of identifying and correcting observations that the observer has associated with the wrong date on their B-91 form. Known as date shifting, such errors can lead to temporal incongruities when the data from an uncorrected, time-shifted station is compared with stations that do not time-shift. The second temporal issue is that COOP stations may not be recording their observations at the time-of-day at which they are supposed to. Because diurnal observations can differ significantly between stations that take morning readings and those that observe in the afternoon, COOP stations must be processed according to their observation time. If their observation time is not correct, the data may be incorrectly compared with those of other stations. Unlike the B-91 form, a Web-based form can associate an observation time stamp with every submission and allow the data to be processed accordingly.

Lastly, it should be noted that, while the paper B-91 forms have in the past provided a convenient original record of the COOP observations for a station over a given month, the forms are not officially certified by the United States government. Only the data that reside in the NCDC archives are officially certified. Thus, the process of collecting, processing, and archiving COOP data can be made completely paperless. However, it is recognized that many within the climate community, COOP observers included, may wish to continue to retain paper copies of the monthly observations at a particular station. A Web-based COOP interface such as WxCoder III can easily facilitate the production of a digital, printable form, which contains all the same information currently contained in a B-91 form, and which appears in the same format.

5. CONCLUSION

Given the digital nature of the Web-based COOP submission process, it is clear that not only can NCDC move toward a paperless data ingest process, but that doing so can dramatically improve both data quality and access times. When fully implemented, it is possible that high quality COOP data from a majority of the network could be made available to the climate community within a day of observation.

NOAA's commitment to excellence in climate services will be strengthened by coordinated movement to a paperless environment that affords routine near real-time data ingest from the COOP network. NCDC, NWS and the RCCs will work diligently over the next year to begin the transition to WxCoder III, including the establishment of new protocols for data ingest at NCDC.

6. REFERENCES

Colorado Climate Center (Colorado State University), 2006: http://www.cocorahs.org/

Environment Canada, 2006: *COOLTAP User's Manual*. Downsview, Ontario, 10 pp.

National Research Council, 1998: *Future of the National Weather Service Cooperative Observer Network*. Washington: National Academies Press, 78 pp.

National Weather Service, 2003: *WxCoder II User's Manual (Version 1.0)*. Kansas City: NWS Central Region, 13 pp.

NERON (NOAA's Envrionmental Real-Time Observation Network), 2006: http://www.isos.noaa.gov/