8.3 HOURLY PRECIPITATION DATA PROCESSING CHANGES AT NCDC

Anna M. Wilson *
STG, Inc., Asheville, North Carolina

Stuart Hinson, Daniel J. Manns, Ron Ray, and J.H. Lawrimore
NOAA National Climatic Data Center, Asheville, North Carolina

1. Introduction

NOAA's National Climatic Data Center (NCDC) has been processing Hourly Precipitation Data (HPD) from a punched paper tape media since the 1960's, when the National Weather Service (NWS) deployed a network of more than 2000 Fischer & Porter (F&P) rain gauges. In recent years, the network has been negatively impacted by increasing maintenance costs, outdated technology, and loss of support from the original equipment manufacturer (U.S. Department of Commerce, 2005). The NWS is now in the process of upgrading the data collection mechanism on the existing F&P gauge network. These F&P Upgrade (FPU) and F&P Rebuild (FPR) units replace the punch block system employed since the network's inception with dataloggers that digitally record 15-minute precipitation totals. At the end of the month, a technician downloads the 15-minute precipitation totals to a data key and ftp's the data to NCDC for processing and archive.

A subset of approximately 200 F&P gauges also have equipment that makes data transmission via satellite possible. The telemetry capability of these stations is interconnected with the punch block technology. These stations will not receive an upgrade from punch block to FPR, but will rely solely on telemetry for data transmission. Data are transmitted on an hourly basis and made available through the NWS Hydrometeorological Automated Data System (HADS). These data are subsequently ingested at NCDC and processed along with data from the punch paper and FPU/FPR sources. A much larger number of Data Collection Platforms (DCPs) other than NWS F&P (e.g., U.S. Geological Survey (USGS) and U.S. Army Corps of Engineers) provide hourly precipitation data via HADS. Although data from other networks are not included in an official NCDC hourly precipitation dataset, the additional data provide a source for improving the overall quality control effort for F&P data. For example, more robust spatial consistency checks will be possible.

For the past several decades, NCDC has relied on a system that was heavily dependent on human interaction to move F&P HPD through the process of acquisition, digitization, quality control, and archive. The NWS effort to transition from punch paper recording to digital data collection via the FPU/FPR upgrades provides a prime opportunity for NCDC to upgrade its quality control and processing system to one that is more automated and that integrates other data resources to improve the overall quality control effort.

The new software will leverage the improved technology into a robust and fully automated processing system, which will take advantage of the increased density of data sources. This system is targeted for implementation by the time the last FPU unit is deployed in 2012. This paper describes the changes in F&P data processing, including benefits to users and quantifiable improvements over the existing system.

2. Processing History

2.1 *HPD History*

HPD have been recorded since 1948, when data were keyed by meteorological technicians onto punched cards for storage and archiving by the regional Weather Records Processing Centers. The task moved to NCDC when it was established in September 1951. In the late 1960s, the HPD data archive was transferred from punched cards to magnetic tape. This data file was converted to the element file structure in 1984.

The HPD network originally consisted of several different weighing rain gauge instruments. F&P gauges, with automated readout recorded on paper tape, were phased into the network in the early 1960s, and they became the primary recording

Figure 1. The inside of a Fischer & Porter gauge using punched-paper tape.

* *Corresponding author address:* Anna M. Wilson, NOAA National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801; email: Anna.Wilson@noaa.gov.

Figure 2. A MITRON reader evaluating punched-paper tape.

instrument. In 1996, there were approximately 2400 F&P gauges in operation. The F&P gauge mechanically converts the weight of liquid in a bucket every 15 minutes to a punched paper tape, shown in Figure 1. Observations are processed by running the tapes across a MITRON punched-paper tape reader (Figure 2) which converts the data on the tape into a digital precipitation record.

Data before 1984 were processed through a gross value check only. Beginning in January 1984, the F&P data were processed through a completely revised system which produced the element structure database file. This system was further enhanced in January 1996, when a graphical user interface and other features were added to aid quality control technicians in identifying and correcting measurement errors (NCDC, 2003).

2.2 *HADS History*

Precipitation data collected and transmitted via the NWS HADS network has served the hydrologic community since 1996. The network, operated by the Office of Hydrologic Development (OHD), receives data from various agencies, such as the U.S. Geological Survey (USGS), the U.S.-Department of Agriculture (USDA) Forest Service and the U.S. Army

Figure 3. The HADS station network. (http://www.weather.gov/ohd/hads)

Corps of Engineers. The different sources that comprise the HADS network make it highly variable in terms of regional spatial density, number of reporting stations, and instrumentation. There are about 6700 stations in the U.S. and its territories that report precipitation data through the HADS network (Figure 3). Approximately 200 of the HADS stations are also part of the F&P gauge network. HADS data are transmitted using Geostationary Operational Environmental Satellites (GOES) telemetry.

The HADS data reports are encoded into Standard Hydrometeorological Exchange Format (SHEF) for transmission to the user community. Near real time data are available for up to one week on the NWS/OHD website. These data are used operationally in flood forecasting and to produce the multi-sensor precipitation estimate. NCDC began processing and archiving HADS data in May 2005. Data from 2002 to the present are available on NCDC's website.

3. Existing System Overview

3.1 *HPD - Data Ingest Procedures*

The F&P network of gauges is the primary source of data for NCDC's hourly precipitation dataset (DSI-3240) and 15-minute precipitation dataset (DSI-3260). These datasets also include data from about 250 ASOS (Automated Surface Observation Systems) stations and 30 Universal rain gauge stations.

Until all eligible F&P gauges are upgraded with FPU/FPR equipment, paper tapes and Universal charts will continue to be sent to a contractor in Kentucky for digitization. The paper tapes are translated by the MITRON reader, while charts are read by hand and keyed directly into a precipitation file by the operator.

The FPU/FPR data are sent to NCDC from regional NWS offices after data are received from all of the region's observers. These data are in a Comma Separated Values (CSV) format. In the intermediary until a new processing system is established, files from the FPU/FPR stations are reformatted to conform with the files created from the paper tapes, which enables processing by the existing software. In addition, the operators at NCDC and in Kentucky receive a 79-ID form for any stations having a paper tape with data that are unlikely to be extracted by the MITRON, or when a station becomes an FPU/FPR station in the middle of the month. Data from these forms are entered directly into the monthly precipitation file. In the first case, the form is entered in Kentucky; in the second, it is entered at NCDC.

3.2 *HPD - Data Processing Procedures*

The current processing system is based on methods developed in 1984, which was the first time automated processing steps were incorporated into the ingest to archive pathway. Algorithms, developed using an objective pattern-matching method, remove non-precipitation effects from the data. These algorithms primarily address spikes and oscillations. Causes include the following: emptying the precipitation bucket, addition of oil or antifreeze, punch failure, punch sensitivity near the gauge thresholds, and others. After the implementation of the new algorithms, it was found that 95% of the paper tapes were adequately corrected for these types of errors (Phillips 1985).

Beginning in 1995, there was an intensive effort to enhance the existing system. The overhauled system became operational in January 1996. The new interactive quality control system introduced computer editing procedures. Specific improvements included visual aids, such as topographic maps incorporating data from nearby stations, and new algorithms, which reduced the number of stations reviewed by technicians by 50%. These algorithms addressed meteorological reasons for a misleading gauge report, such as evaporation and snow melt. When the new software determined that a decrease in the punch measurement was the result of evaporation, it began counting precipitation increases again at that point, instead of waiting until the gauge weight returned to its previous peak. New checks helped the technician identify snow melt, and a new flag was created to describe the event more precisely.

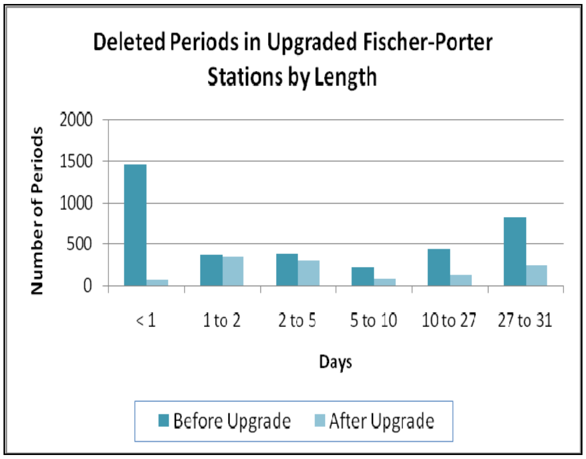

Figure 4. This graph shows the number of deleted periods in 245 stations before and after upgrade.

The technician's prior response to a report caused by snow melt was to accumulate the value over the likely time period of the precipitation, using neighboring stations for guidance (Hammer and Reek, 1997). In the new system, the technician instead added a flag. This way, the original data value and its level of uncertainty were represented accurately to the user.

With the transition from paper tapes to digital recording via the FPU/FPR, NCDC has begun to evaluate changes in data quality resulting from the new technology. Results show that the upgrade significantly reduces flagged periods in the data. For example, when data from 245 FPU stations are compared before and after the upgrade, deletions decrease, on average, from 15% of the data points per station over the month to 5%. Figure 4 shows these decreases in terms of the number and length of deleted periods. Suspicious values and questionable observation times also decrease, by 25% and 48% respectively. An evaluation of the number of missing observations showed an unexpected increase over the number present each month before the upgrade took place. Further review found that the problem resulted from a lag in updating the station's metadata to reflect the fact that the upgrade had taken place rather than any physical problems experienced by the gauge or recording device. This highlights the critical importance of timely and accurate updates to metadata collection and access systems.

3.3 *HPD - Data Output*

Hourly and 15-minute precipitation data are available in DSI-3240 and DSI-3260 respectively. The data are also available in a monthly publication that can be obtained in hard copy or on NCDC's website. During the 1996 rehabilitation project, 15 minute raw gauge values, which are the measurements of the total liquid in the gauge at the time the punch is recorded, began to be retained in DSI-3260, along with flags assigned to them during the automated quality control processes. DSI-3260, DSI-3240, and the HPD publication all contain edited precipitation values and flags. In the new processing system, raw data values will continue to be retained and available to the user, along with the QC algorithms used and the resulting edited values.

3.4 *HADS – Existing System*

Precipitation data distributed via HADS originate from a variety of rain gauge types depending on the contributing agency. Generally they are either weighing or tipping bucket gauges. However, all gauges in the HADS network relay information through GOES DCPs. Data are received from the GOES satellites at the Wallops Island Virginia Flight Facility on a continuous basis. Gauges report at varying frequencies, from 5 to 360 minute intervals. The variety in reporting time extends to stations that report hourly totals at times other than the top of the hour, which can cause confusion when comparing the hourly data. After receipt, data are converted to SHEF format and passed through the HADS network to NWS River Forecast Centers (RFCs), Weather Forecast Offices (WFOs), NCDC, and other users.

Using data collected via HADS, the NWS OHD implements four levels of semi-automated quality control to support NWS hydrologic operations: gross error checks, outliers based upon season and geography, spatial quality control checks against neighboring sensors, and expert judgment (Kondragunta and Shrestha 2006). Quality control is performed throughout the HADS data flow. Checks begin during data collection, when suspicious values are flagged, and end at the database, where values that failed checks can be excluded from an application.

At NCDC HADS data are received once per day. The reprocessed HADS data available on the NCDC website go through several automated quality control checks. These checks revise missing data values when possible and remove spikes and noise from the dataset (Kim et al. 2009). The reprocessed data reduce the number of missing values from 5% to 1%, and increase the number of top of the hour observations from 50% to 85%. These reprocessed data are available on NCDC's website and updated once daily as data are received (http://www.ncdc.noaa.gov/hads/).

4. New Quality Control Methods

NCDC has initiated an effort to upgrade its quality control and processing systems for hourly precipitation data based on the paradigm established in Durre et al. 2008. This model was most recently applied in the development of quality assurance processes for NCDC's Global Historical Climatology Network-Daily (GHCN-D) dataset (Durre et al. 2009). The strategy involves complete automation in the form of a robust and reliable quality control system, in which data are analyzed consistently and objectively. Manual intervention is used extensively prior to the implementation of the quality control algorithms to ensure the validity of thresholds and logic in the system's decision making. This differs from the traditional semi-automated process, where the decisions made by automated procedures are manually evaluated as part of the operational quality control process and sometimes overridden. Thorough documentation of the system's performance is required, including an empirical assessment of false-positive and flag rates, information on types of errors removed and detected, as well as conditions under which errors might remain. Documentation on the processes and thresholds applied in the quality control process should be available to users to aid them in making an informed decision about how to appropriately apply the data.

Advantages to this method over traditional methods involving manual intervention include the removal of the subjective component intrinsic to any process with a human interface. This method also provides a consistent set of quality control checks throughout the period of record, instead of antecedent practices that introduce new quality measures at various times throughout the period of record. Most importantly, the ability to process the entire period of record makes it possible to apply quality control retrospectively as new methods are developed and to do so in a consistent manner throughout the life of the data.

Although improvements have been implemented to the F&P HPD processing system since its inception, it remains heavily reliant on manual intervention. The automated component to the system filters and smoothes obvious oscillations and spikes as defined by the objective pattern-matching algorithms, but the rate of stations flagged for manual inspection is high. Figure 5 indicates the

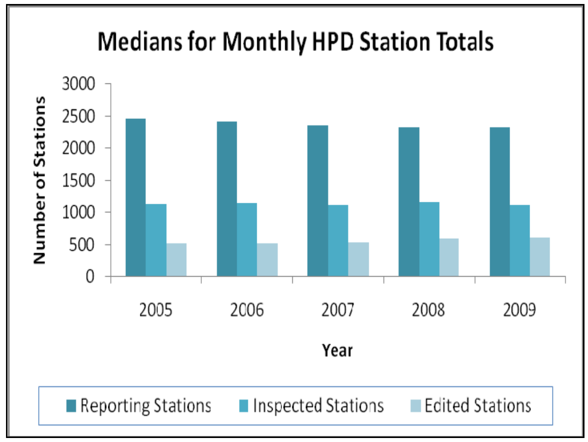

Figure 5. This graph shows the median number of stations reporting, manually inspected, and edited for the month during each yearly period.

median numbers of stations reporting, inspected and edited over yearly periods from January 2005 – April 2009. During this entire period, an average of 47% of reporting stations were flagged for review. Approximately 45% of the flagged stations required modification during inspection.

While the F&P HPD system has performed well over the past 14 years, recent advances in quality control techniques and observing technologies provide an opportunity to upgrade the processing system and create an automated and robust quality control system to produce consistent, reliable datasets with better timeliness. The extensive documentation on the existing system, input from technicians currently working on producing the monthly HPD archive and publication, inclusion of additional data sources and the paradigm now being applied to other datasets at NCDC will all contribute to the new system's development.

The comprehensive revamping of the F&P HPD process will introduce other data streams into the system to improve quality control through the introduction of new spatial and temporal consistency checks. Currently, the system uses Cooperative Observer (COOP) Summary of the Day (SOD) data for spatial checks. SOD data have a daily resolution and observation times that vary, and thus provide limited usefulness for the quality control of hourly data. By leveraging additional sources of hourly precipitation data from HADS, it will be possible to augment existing automated quality control processes to reduce or eliminate human interaction while improving the overall quality control process. Other sources of data, such as lightning data for convective screening, and radar and satellite data for detection of stuck gauges (Kondragunta and Shrestha, 2006) will also be evaluated for inclusion in the new system.

5. Summary

The new F&P HPD quality control system will be completely automated, rendering it fully objective and reproducible. The system will use precipitation data, from networks such as HADS, and data from other sensors such as lightning and radar for high resolution spatial and temporal analysis. This represents an improvement in terms of manual handling of the data, which will be eliminated after system validation, and in terms of the spatial analysis, which will be increasingly meaningful through use of more relevant precipitation networks. Older data can be reprocessed to reflect this increased reliability through the historical record. Raw data values will always be retained to ensure the ability to reprocess data as advances are made and adjustments to the system required.

The new system will increase the information available to the user and make quality controlled data available in a more timely manner. The intensive manual effort developing the new system, including validation studies and quantitative evaluation of the system's improvement, will result in a reliable and thoroughly documented dataset. All data adjustments made by the software will be recorded. Familiarity with the data's production will ensure that the user understands any possible issues with the data, and will help guide applications.

Figure 6. An example of a new precipitation recorder, which will replace the paper-tape assembly in Fischer & Porter rain gauges (U.S. Department of Commerce 2009).

This is expected to be a multi-year effort with a targeted completion data corresponding to the completed upgrade of all eligible F&P stations. Over the next several years each eligible F&P station will be equipped with a modern sensor and a data logger, such as the precipitation recorder in Figure 6. Lower rates of flagged data have already been documented at stations that have received a hardware upgrade, and advances associated with development of the new quality control and processing system is expected to provide even greater improvements in quality and data availability.

6. References

Durre, I., M. J. Menne, and R. S. Vose, 2008: Strategies for evaluating quality-control procedures. *J. Appl. Meteor. Climatol.,* **47**, 1785-1791.

Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston and R. S. Vose, 2009: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.,* submitted.

Hammer, G. R. and Reek, T., 1997: The processing of recording rain gage data at the National Climatic Data Center. *Proc. 13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 223-226.

Kim, D., B. Nelson, and D. J. Seo, 2009: Characteristics of reprocessed Hydrometeorological Automated Data System (HADS) hourly precipitation data. *Wea. Forecasting,* **24**, 1287-1296.

Kondragunta, C. R. and Shrestha, K., 2006: Automated real-time operational rain gauge quality-control tools in NWS hydrologic operations. *Extended Abstracts, 20th Conf. on Hydrology,* Atlanta, GA, Amer. Meteor. Soc., P2.4.

NCDC, 2003: Data documentation for data set 3240 (DSI-3240), National Climatic Data Center, 151 Patton Ave, Asheville, NC 28801-5001, USA.

Phillips, C. S., 1985: An objective method for minimizing non-precipitation effects in precipitation data from punched paper tape. *Proc. International Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, Los Angeles, CA, Amer. Meteor. Soc., 178-182.

U.S. Department of Commerce, 2005: *Cooperative Observer Program Product Improvement Implementation Plan [Addendum I] for Fischer & Porter Sensor Upgrade*. NOAA, NWS, 29 pp.

U.S. Department of Commerce, 2009: *NWSREP's Fischer-Porter Rebuild (FPR-D) Operations Manual*. NOAA, NWS, 87 pp.