Algorithm Theoretical Basis Document

COOP-Hourly Precipitation Data (HPD) Version 2

C-HPD/ATBD
Version 2.0.0 / October 23, 2018

RESPONSIBILITY

Prepared By:	Jay Lawrimore	Chief, DS	NCEI/CWC/CSB
	Signature		Date
Reviewed By:	Jeff Privette	Deputy, CWC	NCEI/CWC
	Signature		Date
Reviewed By:	<Name>	<Title>	<Organization>
	Signature		Date
Approved By:	Jeff Privette	Deputy Dir., CWC	NCEI
	Signature		Date
Approved By:	<Name>	<Title>	<Organization>
	Signature		Date

REVISION HISTORY

Version	Description	Revised Sections	Date
2.0.0.beta	Initial submission (Beta)	New Document	02/01/2017
2.0.0	Update for Operational Release	Operational Release	10/23/2018

ACRONYMS AND ABBREVIATIONS

Acronym or Abbreviation	Meaning
COOP	U.S. Cooperative Observers Network
GHCN-D	Global Historical Climatology Network-Daily
LST	Local Standard Time
NCEI	National Centers for Environmental Information
NWS	National Weather Service
NOAA	National Oceanic and Atmospheric Administration
QC	Quality Control
UTC	Coordinated Universal Time

1. Introduction

1.1 Purpose

The purpose of this document is to describe the algorithms, software, and datasets submitted to the National Centers for Environmental Information (NCEI) by the Dataset Section (DS), Climate Science Branch (CSB), Center for Weather and Climate (CWC) that are used to produce the Hourly Precipitation Dataset version 2 dataset. Data are collected from the NOAA/NWS Fischer-Porter in situ network of precipitation gauges. The data within each Weather Forecast Office (WFO) area of responsibility are collected via digital download, collated at the WFO office, and delivered to NCEI's ftp site on a monthly or other periodic basis. They are subjected to a series of processes that combine the data with the historical record (DSI-3240; i.e., version 1), converted from 15-minute gauge depth values to precipitation increment, summed into hourly totals, and then subjected to additional quality control before being output into datasets for customer access and permanent archive. Customers are primarily external to NCEI. The actual algorithms are defined by the computer programs (code) that accompany this document, and thus the intent here is to provide a guide to understanding those algorithms, the software, and datasets, from both a scientific perspective and in order to assist a software engineer performing an evaluation of the production and update process.

1.2 Document Maintenance

This document will be maintained in a manner consistent with version control practices for NCEI. When a new version of the global land monthly process is developed, this document will be reviewed and edited as necessary to ensure it remains consistent with the current operational version.

2. Dataset Overview

2.1 Dataset Descriptions

The NOAA/NWS Fischer-Porter (F&P) network, consists of approximately 1900 stations that comprise a subset of the NWS Cooperative Observers Program (COOP) network. It is the source of observations for NCEI's Hourly Precipitation Dataset (HPD), known historically as DSI-3240 (NCDC 2003). The new dataset described herein combines the legacy data from DSI-3240 with a new source of data collected from the same Fischer-Porter gauges, but which were upgraded to digital recording beginning in the mid-2000s. As with the legacy 3240 dataset, this new dataset provides observations of hourly precipitation amounts from the 1940s to the present.

In addition to supporting studies on climate variability and change, the lengthy record of this network combined with its relatively high spatial density provides an important source of data for hydrometeorological applications, supporting water resource decision makers, engineering, and weather forecasting including flood warnings.

Until recently the F&P network operated through mechanical conversion of the weight of liquid collected in the instrument bucket to a punched paper tape record of the depth of precipitation every 15 minutes (Phillips 1985). In 2004 the NWS began upgrading the F&P stations by replacing the paper tape with a digital recorder at most sites (US Department of Commerce 2005, 2009). The upgrade designated as the Fischer-Porter Upgrade (FPU) when it was first initiated in 2004, and later as the Fischer-Porter Retrofit (FPR), aimed to improve the quality and completeness of the precipitation observations while reducing maintenance costs. Instead of recording on a paper tape which was subject to tearing, deterioration, and being expended before a technician's monthly visit, the upgrade implemented digital recording on a datalogger. The design still requires monthly site visits for data downloads to a data key (thumb drive), but problems associated with paper records are no longer a concern.

Accompanying this transition to digital recording, NCEI designed a new data acquisition, integration, and quality control process. This involved a change from a largely manual review and edit process to a fully automated system that removes the subjectivity that was previously a necessary part of quality control and processing. The conversion from paper to digital recording along with development of an objective quality control process is aimed at greatly improving the quality of the HPD data. A description of this automated processing and quality control system is provided in the following sections which are organized as follows: section 2 describes the data sources, section 3 the data acquisition process including conversion of gauge depths to 15-minute precipitation totals and the quality control process for hourly precipitation totals, testing and validation procedures are described in section 4, practical considerations including operational processing are included in section 5, and references are provided in section 6.

2.2 Data Sources

The HPD network has operated since the middle of the 20th century, providing 15-minute precipitation observations and hourly totals for approximately 1900 stations. It originally consisted of several different weighing rain gauge instruments before F&P gauges, with automated readout recorded on paper tape, were phased into the network in the early 1960s (Wilson 2010). Small changes in the number of stations have occurred throughout the period of record, but the total number of stations in the network has remained fairly steady. The locations of the stations in the HPD network are shown in Figure 1. The various station configurations are shown in the legend and discussed in section 2.4.

2.3 Legacy punched-paper data

The historical record of 15-minute and hourly observations in the HPD dataset consists primarily of data from F&P punched-paper recorders. Until the conversion to digital recording, data collection required removal and replacement of the paper tapes each month by a trained technician at each NWS WFO. The tapes were packaged and sent to NCEI or one of its contract facilities for processing. Precipitation data were extracted by running the tapes through a MITRON punched-paper tape reader which converted the observation into a digital precipitation record of gauge depth at 15-minute increments. The gauge depths were subsequently converted to 15-minute precipitation amounts and summed into hourly totals as part of operational processing at NCEI. The punched-paper recorder provided 15-minute amounts in increments of 0.10”.

Stewardship and access to the hourly and 15-minute observations was provided through datasets designated as DSI-3240 and DSI-3260 datasets, respectively. The hourly observations in the DSI-3240 dataset were augmented by hourly precipitation totals from approximately 270 stations that were also part of NCEI’s Local Climatological Data publication as discussion in section 2.6.

2.4 Digital Recording

The punched-paper recording device provided a half-century record of precipitation across the U.S. However, observations recorded on the punched paper tape were subject to loss due to tearing, deterioration, and the tape being expended before a technician's monthly visit. As aging equipment became more difficult to maintain, and with improved methods for precipitation measurements available, the NWS began upgrade and replacement of the punched paper equipment in 2004.

The upgrades took place in a series of deployments beginning with approximately 250 stations designated as F&P Upgrade (FPU) equipment. This was followed by deployment of F&P Retrofit (FPR) equipment designated FPR-D and FPR-E. Almost all FPU equipped stations were replaced by FPR-D/E equipment before the end of the modernization effort in 2014. The new design still requires monthly site visits for data downloads to a data key (thumb drive), but problems associated with paper records are no longer a concern. The last of 1855 stations were converted to digital recording in 2014.

The various designations reflect different vendor-specific equipment. The FPR-D uses a load cell technology to measure gauge depth, while FPR-E equipped stations use a strain gauge. All equipment provides a digital record of 15-minute gauge depths recorded on dataloggers at each site. Precipitation is recorded in increments of 0.01"; however, the accuracy of the instrument is 0.10" and the calibration tolerance is 0.24". Thus while precipitation measurements are made to 0.01" it is important to communicate to the user community the limitations in recording very light precipitation amounts.

Collection and dissemination of the data is made possible with monthly site visits by NWS Weather Forecast Office (WFO) personnel or volunteers who download the data to a portable memory stick. The data are subsequently transferred to a computer at each WFO for the stations in their respective Forecast Area (FA). In many cases the files are e-mailed from the local observer to the WFO. The data files for all stations in the WFO FA are subsequently bundled into a Windows Zip file by NWS personnel and transferred to NCEI via ftp.

Each Zip file contains a set of files containing 15-minute gauge depths for each F&P station. As these files are transferred to NCEI's ftp site, they are subsequently processed through a fully automated system. This system includes ingest, archive of each zip file as it is received, and subsequent parsing and downstream processing. An automated review of each station file is made to identify quality problems associated with filename conventions and data corruption so that feedback can be provided to the NWS WFOs. This is described in more detail in section 3.1.

2.5 Other F&P stations (HADS, 79-IDs)

Approximately 150 F&P stations were not designated by the NWS for inclusion in the digital recording upgrade project (and they are not included in this release of the C-HPD dataset). However, these stations continue to operate in part to support precipitation forecast and warning activities for the NWS, the US Army Corps of Engineers, and the US Geological Survey. The data for these stations are collected in the methods described below. Note that these data will need to be included in future releases of this C-HPD dataset.

2.5.1 HADS

Data (15-minute totals) for 60 to 70 stations are transmitted hourly via satellite or telephone telemetry and collected as part of the NWS Hydrometeorologial Acquisition Data System (HADS). Due to outages that periodically occur in the transmission, collection, and transfer of data through the telemetry process, data completeness is lower for these stations than it is for the digital recording that takes place at the other F&P stations. However, these data are available to NCEI via HADS and can be incorporated into the new HPD dataset in a future version.

2.5.2 NWS Form 79-IDs

Hourly precipitation totals for an additional 70 to 80 stations that were not upgraded are converted from the punched paper tape record at some WFO offices and sent to NCEI via electronic NWS Form 79-IDs. These forms provide hourly precipitation totals instead of the 15-minute gauge depth values that are available via the stations operating with digital recording. The data from this source are being collected and archived at NCEI and can be incorporated into the HPD dataset in a future version.

2.6 Local Climatological Data

For many years the F&P data in the DSI-3240 (HPD) dataset were augmented by approximately 270 stations operating a major airports and NWS offices. These stations consisted primarily of manual observations of hourly precipitation totals until the transition to the Automated Surface Observing System (ASOS) took place in the 1990s. The hourly precipitation totals from these stations were processed and quality controlled as part of NCEI's Local Climatological Data (LCD) program, but they also were merged with hourly precipitation totals from the F&P network to improve the spatial coverage of hourly observations in the legacy DSI-3240 dataset. This was especially helpful in sparsely populated areas of the central and western U.S. The integration was necessary for providing the user community with hourly data of sufficient spatial coverage. But with modern web-based methods of data access, it is no longer necessary to include this subset of LCD stations within the HPD data set. Thus these stations are excluded from HPD version 2 processing.

Algorithm Description

3. Processing Outline

Please refer to the Level 0 and Level 1 flow diagrams in Figure 2 and 3 as an aid in understanding the processes described in sections 3 and 4.

As introduced in section 2, each NWS WFO is responsible for placing data for each F&P station in their Forecast Area (FA) onto NCEI's ftp server. This is accomplished by consolidation of the station files into a zip'd file that contains data from the FA. Data are typically transferred once a month, following collection of data for the prior month. However, more than one month can pass between data transfers.

3.1 Data Ingest and Integrity Check

Uploads from each of the 122 WFO offices to NCEI's ftp server occur on a continuing basis. As the data arrive they are subjected to a once-a-day archive process which preserves the data in its original form. In addition, the data are picked up from the ftp site as part of the ingest process. The processing of these data begin with unzipping of each set of station files, and a check is performed to confirm the files are successfully unzipped. If not successful the files are quarantined and automated notification to NCEI staff occurs via e-mail so that the issue can be communicated with the NWS. Similar checks are performed further downstream and notifications sent for checks that fail.

For files that are successfully unzip'd, a check of the filename convention for each station is made to ensure consistency with prescribed requirements. A finding of a valid filename results in the file being placed into a directory specific to the type of F&P gauge. The dates within the file are subsequently checked to confirm their validity. A finding of a valid date range is followed by additional checks for data formatting which is followed by removal of any data that were duplicated within the current or previous transfer. Any data or metadata other than precipitation observations from the F&P gauges also are removed. Because the process of data collection and transfer requires a large amount of manual file manipulations at each WFO, it is not uncommon for files unrelated to the F&P gauges to be included in the zip'd files transferred to NCEI. The status of processing for each set of station files is then written to a log file that is available online at http://www1.ncdc.noaa.gov/pub/data/hpd/inv/ . This provides information that each NWS WFO can use to confirm that their data were successfully processed at NCEI or to determine the source of any problem (e.g., poorly formatted data) that needs to be resolved.

Data successfully processed to this point are loaded into an Oracle database table for storage of the observed gauge values. This automated process occurs on an hourly basis as new data arrive on the NCEI ftp server. Managing the data in this way including retention of the original zip'd data file allows for full reprocessing from the original observations if needed.

Data that reach the Oracle database are subsequently extracted and loaded into text files on the hpd 3-tier server for further processing.

3.2 15-minute Gauge Depth Processing

Observations are recorded as the depth of precipitation in the gauge every 15 minutes. At first glance it may seem a simple problem to subtract one 15-minute gauge depth value from the previous gauge depth to calculate the amount of precipitation that fell in the previous 15 minutes. However, there are a number of complicating factors that must be addressed. These include missing observations, high frequency low and high amplitude noise, malfunctioning observations that may be due to factors such as inadequate gauge maintenance, radio interference (RF), and small earthquakes that are now prevalent in some parts of the country. Evaporation is also a complicating factor that must be dealt with in converting from gauge depth to precipitation amount. The following algorithms (Figure 4) were originally developed for the NWS HADS precipitation data (Kim and Seo 2009) and modified for the HPD network.

3.2.1 Pre-Process Gauge Depths

The first step in the process (subroutine Prepro_HPD.R) involves a basic review of the gauge depth data and reported time of observation. Observations reporting near (within 1-minute) but not precisely at the top and bottom of each hour and on the 15-minute increment between hours, are assigned to the appropriate incremental time period. In addition, observations at off-times are excluded.

3.2.2 Missing Depths Filled & Bucket Drain Check

It is not uncommon to find instances of one or more missing 15-minute gauge depth measurement in a monthly HPD file. When producing the monthly series of incremental changes in gauge depth it is often possible to replace a period of missing observations with estimated values if the period of missing data is less than one day. In this step (subroutine Remove_NAs.R), the gauge depths at the start and end of each missing period are compared to determine if infilling is possible. In instances when the gauge depth values are identical before and after the missing period, the missing period of 15-minute values is filled with the adjacent gauge depth. If the difference between the beginning and end of the missing period is less than or equal to 0.51 mm (0.02 in.), the missing values are replaced with the average of the gauge depth values on each side of the gap. Otherwise the values remain missing; this occurs in only 1055 out of 64162 data-months from 2006 through 2013.

3.2.3 Empirical High Frequency Noise Removal

High frequency low amplitude noise is a common issue with the Fischer-Porter instruments. This can occur due to factors associated with equipment age and the level of maintenance for a station as well as natural and manmade influences such as fluctuations associated with small earthquakes, and the passage of nearby rail traffic. In this step (Step1_QC.R), small negative and positive changes (less than +/- 0.03”) in gauge depth that occur in the absence of precipitation are removed by locating offsetting negative and positive oscillations with equal magnitudes within 3-hour periods. When a small reduction in the gauge depth is followed by an equal rise in the gauge depth within the same 3-hour period, the negative and positive increments are removed (i.e., the incremental change in gauge depth is set to zero). 3-hour blocks are evaluated from the beginning to the end of the month. If a compensating increase in gauge depth does not occur within the same 3-hour period, the negative increment is held over for the next module.

3.2.4 Malfunctioning Gauge Depth Identification

Large fluctuations unrelated to precipitation of magnitudes greater than 0.05” sometimes occur due to factors such as low voltage power supply, poor gauge maintenance, or electromagnetic interference. These fluctuations can occur over multi-day periods often stretching for one or more weeks. If precipitation (generally light to moderate) occurs during these periods it is generally not possible to accurately determine the amount of precipitation because the increment is less than the amplitude of the noise. Algorithms were developed to identify these periods, to flag observations as invalid when necessary, and to preserve as much of the record of precipitation as possible. In this step (SANITY_qHPD.R), three sets of logical conditions are used to identify such conditions. The checks are performed using the 15-minute gauge depths reported over a full day (0000L to 2359L). If any of the three conditions listed below are met, the observations for the entire day are invalidated.

This module begins by computing the incremental precipitation values for each 15-minute period in the day. Sample statistics are computed from the distribution of ninety-six 15-minute precipitation increments; their mean, standard deviation of trimmed samples (those increments less than 5.1 mm (0.20”)), and the number of negative increments in the day. The slope across all 96 increments of the day is computed (any increment greater than 20.3 mm (0.8”) is excluded from this calculation. The data are then trimmed by removing the 15-minute increments more than 5.1 mm (0.20”) from the mean. From the remaining distribution of 15-minute data, the following three conditions are evaluated. If any of the following conditions are met, the observations for the day are invalidated.

1) Standard deviation of the trimmed data is greater than 0.76 mm (0.03 in.), and the mean is smaller than 0.05 mm (0.002 in.).
2) Standard deviation of the trimmed data is smaller than 0.76 mm (0.03 in.), the maximum increment for the day is no greater than 0.25 mm (0.01 in.), the number of negative increments is greater than 12, and the number of positive increments differs by no more than six from the number of negative increments.
3) The number of negative increments in a day is less than 20, the slope is less than 0.13 mm (0.005 in.) and the maximum increment is greater than 20.3 mm (0.8 in.).

The robustness of such empirical threshold values was tested with hundreds of station-months to determine the optimum threshold values. The thresholds were established to reduce the chance of flagging and removing valid periods when the instrument was performing well, while also retaining the ability to determine the amount of precipitation that fell when the amount exceeded the noise level. Figure 4 provides two examples of an instrument malfunction. One occurs throughout the entire month, although it was possible to determine the amount of precipitation that fell when the amount exceeded the noise level. The days that were invalidated are marked with the letter “F”.

3.2.5 Removing the Effect of Diurnal Fluctuations

Diurnal fluctuations of bucket values on non-precipitating days are omnipresent in almost all F&P gauges with varying degrees of amplitude depending on the level of routine and unscheduled maintenance performed and the climate zone. For each day of a month, this step (Step2_QC.R) determines if small fluctuations in gauge depth are caused by diurnal variations unrelated to precipitation.

This module begins by summing the incremental changes in 15-minute reported gauge depth throughout the month. These are referred to as cumulative precipitation (CP) at any time during the month. The difference between the smallest and largest CP values for each day of the month are determined (i.e., daily amplitude). Then the lowest 10th percentile of these daily amplitudes are computed.

Each day of the month is evaluated against the 10th percentile value to determine if the day was wet (precipitation fell during at least one 15-minute period) or dry (precipitation did not occur). A day is determined to be dry if the difference between the current day and previous day's lowest CP is less than the 10th percentile AND the difference between the current day and previous day's highest CP is less than the 10th percentile. If the current day is identified as a dry day, all 15-minute increments for the day are set to zero. Otherwise the 15-minute incremental values for wet days remain unchanged and are carried over for processing in the subsequent step.

An example of a diurnal cycle is seen in Figure 5 during April 2011 for Crown King, Arizona (COOP ID 022329). The increments of fluctuations in bucket values during those days are manifested as strings of positive values and negative values. In this step the small incremental positive and negative changes are set to zero.

3.2.6 Writing to Internal Format

The final step of the 15-minute gauge depth filtering process is the output of 15-minute precipitation totals (Write_15m.R). These data are then used in subsequent processing.

3.3 Data Merge

Hourly totals are computed from the 15-minute totals. The 6-digit COOP IDs for each station are converted to the respective 11-digit GHCN-D identifier. The hourly precipitation totals are then merged with Period of Record data on a station-by-station basis. The data for each station are then subjected to quality control checks as described in the next section.

3.4 Quality Control of Hourly Precipitation Totals

The hourly precipitation totals are subjected to an automated quality control process based on a paradigm established by Durre et al. (2008) and first applied in development of quality assurance processes for NCEI's Global Historical Climatology Network-Daily

(GHCN-D) data (Durre et al. 2010, Menne et al. 2012). 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 of 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 evolving 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.

3.4.1 Global Extreme Exceedance Check

This check flags any hourly value that exceeds the all-time global hourly extreme of 12.0 inches (305 mm).

3.4.2 Negative Precipitation

Any hourly value less than 0.0 inches is flagged as invalid. Although this check is included on the hourly timescale, the processing as described in section 3.2x typically result in no negative hourly precipitation totals.

3.4.3 State Daily Extreme Exceedance

The sum of the 24-hourly totals in a calendar day is compared to the all-time daily state extreme within the respective state for each station. All hourly values within the day are flagged on a day that any daily total exceeds the all-time state precipitation extreme.

3.4.4 Streak Check

Streaks of identical values are flagged if the streak exceeds the following threshold. In evaluating streaks of precipitation, missing and zero hourly totals are ignored. A streak of 20 or more identical non-zero precipitation values less than or equal to 7.6 mm (0.3”) results in all values being flagged as invalid. Likewise values are flagged for streaks of five or more identical non-zero precipitation values greater than 7.6 mm (0.3”).

3.4.5 Gap Check

The distribution of hourly precipitation totals within a 31-day moving window are computed on a station-by-station basis. This distribution is used to identify outliers in hourly precipitation. A gap in the distribution of hourly totals of non-zero precipitation greater than 1.25” results in flagging of the hourly total on the upper side of the gap.

3.4.6 Climatological Outlier

The same 31-day moving window is used in the climatological outlier check. All hourly totals more than 7 times the 95th percentile of a station’s distribution are flagged as invalid.

4. Testing and Validation

The hourly HPD data were tested and validated against the hourly precipitation data of the US Climate Reference Network stations and summary of the day data from the NWS Cooperative Observers Network (COOP; maintained as part of the GHCN-Daily dataset). Three separate sets of tests were conducted.

1. Spatial comparisons of the frequency of hourly totals exceeding various thresholds on a seasonal basis (HPD digital versus CRN)
2. Comparisons of HPD (and CRN) against COOP data to determine if HPD (CRN) recorded precipitation when days were otherwise determined to be dry (based on COOP)
3. Comparisons of HPD (and CRN) event totals against COOP event totals for 743 precipitation events; for a subset of 20 HPD stations co-located with a COOP station and within 10km of a CRN/RCRN station.

4.1 Frequency of hourly totals exceeding thresholds

The frequency of occurrence of hourly precipitation amounts exceeding three different thresholds were evaluated for HPD and CRN stations (0.30”, 0.10”, 0.04”). This was performed for each season. Four-panel seasonal maps are shown in Figures 6, 7, and 8 for 0.30” (7.6mm), 0.10” (2.5mm), and 0.04” (1.0mm) thresholds, respectively. These show that in all seasons the frequency of the light to heavy precipitation amounts are consistent between the HPD and CRN stations. There are a small number of outliers in the HPD network due to very short periods of record that result in what appear to be precipitation that occurs with a higher frequency than surrounding stations.

4.2 Comparison with CRN and GHCN-Daily on dry days

Due to the potential for random occurrences of noise in the HPD gauge depth signal, there is the potential that precipitation (typically very light) may be incorrectly reported during dry periods. Although the filtering and quality control processes discussed in section 3 are designed to remove noise before calculation of hourly precipitation totals, the following analysis was performed to determine the effectiveness of the filtering and quality control processes.

To determine if precipitation in the HPD network might be reported when precipitation has not fallen, the HPD observations were compared to data from nearby stations in the CRN network and the COOP network. Dry days were identified using the GHCN-Daily dataset during periods when no precipitation was recorded within a 5-day window. The center-day of these periods were analyzed for both the HPD and CRN stations to ensure that time of observation differences in the COOP network would not affect the analysis. A subset of HPD stations within 10km of CRN/RCRN stations was used. There were 192136 hours within the center-day of the GHCN-D identified dry periods for which HPD and CRN data exist (since 2005). Of this sample there were 1872 hours (0.84%) in which precipitation of at least 0.01” (0.25mm) was reported by an HPD station (Table 1). There were 578 hours (0.30%) in which precipitation of at least 0.2mm was reported by a CRN station.

Although HPD stations reported precipitation on approximately 1% of the “dry” hours, almost all of these were hourly totals less than 0.03”. Given the design limitations of the HPD sensors (accuracy to 0.10” as discussed in section 2.4) and their sensitivity to other factors that affect gauge depth recordings, it is not surprising that there would be some differences between HPD and GHCN-D. In addition it is likely that some of the differences between HPD and GHCN-D were due to the absence of precipitation reports in GHCN-D on days when rain actually fell (e.g., due to evaporation of light precipitation in the 8-inch gauge between daily observations or due to observer error). On more than half of the 578 hours that precipitation was observed at a CRN station, the HPD station also recorded precipitation, suggesting that inaccurate COOP observations are at least partly the source of the observed differences.

4.3 Comparisons during precipitation events

To evaluate the performance of the HPD network during periods of precipitation, events were identified for days in which there were one or more days of measured precipitation in GHCN-D between two or more consecutive dry days. The analysis was performed for 743 rain events meeting this criteria and for which the GHCN-D precipitation total exceeded 25mm (where CRN is within 10km of an HPD and COOP station). The hourly HPD totals (and CRN hourly totals) were summed into daily totals for comparison with GHCN-D.

In general there was a fairly symmetric spread of differences between HPD and GHCN-D and also between CRN and GHCN-D (Figure 9). There were a greater number of instances in which HPD daily totals were less than GHCN-D (HPD 408 less/323 more/11 tie). The same was true for CRN to GHCN-D comparisons (CRN 460 less/273 more/9 tie).

The scatterplot (Figure 10) also shows strong correlation between HPD (and CRN) with GHCN-D totals for most precipitation events. Regressions show an underestimation (lesser daily totals) compared with GHCN-D as also noted in the paragraph above for both the HPD and CRN networks.

The results of each of the three analyses provide confidence in the quality of the HPD network and the new automated system. The HPD to GHCN-D comparisons and their similarity to the CRN to GHCN-D comparisons provides further confidence in the quality of the version 2 hourly precipitation data from the HPD network and the new automated processing system.

5. Practical Considerations

5.1 Processing Environment and Resources

HPD development is performed on the virtual server hpd-dev. hpd-test and hpd-prod servers were subsequently established for testing and production, respectively. The test environment should be identical to the production server and provides the means for a final checkout before the system is elevated to operations on the production server.

The process runs optimally with 72 GB of RAM and 12 Intel Xeon 5570 2.93 GHz CPU's. No more than 100 GB of storage is required.

The HPD software is a combination of Fortran95, R, IDL, and bash scripts. All software is maintained in the NCEI-NC subversion repository. The system was designed to take advantage of the multi-processor configuration; simultaneous processing enables routine updates to be completed in less than one day. A period of record update and full quality control requires three to four days to complete.

5.2 Change Management

The dataset is version controlled using a three-digit numbering system (x.y.z). The three-digit versioning tracks changes resulting from minor bug fixes up through major structural enhancements. Each change is documented in a manner consistent with the magnitude of the enhancement. Minor bug fixes are recorded in an online status file, moderate changes are described in a Technical Note, and major updates may be communicated in the peer-reviewed literature. Updates to this document will be made whenever there are changes that result in a moderate or major upgrade. An online status file will be updated for minor to major updates.

The file naming structure is hpd.vX.Y.Z.YYYYMMDD where

1. X is incremented when there is a major change to the dataset such as implementation of a new bias correction algorithm or new quality control system.

2. Y is incremented when there are one or more significant changes to the dataset such as the implementation of a single new quality control algorithm or the addition, correction, or removal of a large number of stations.

3. Z is incremented when any minor change is made. These can include minor bug fixes, correction of minor data errors, minor changes to bias correction or quality control processes, and small additions of new station data.

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 (2010), Comprehensive automated quality assurance of daily surface observations, *J. Appl. Meteorol. Climatol.*, **8**, 1615–1633.

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.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston (2012), An overview of the Global Historical Climatology Network-Daily Database, *JAOT*, **29**, 897–910, doi:10.1175/JTECH-D-11-00103.1.

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.

Wilson, A.M., S. Hinson, D.J. Manns, R. Ray, J.H. Lawrimore, 2010: Hourly Precipitation Data Processing Changes at NCDC. *Proc. 15th Symposium on Meteorological Observation and Instrumentation, Atlanta, GA*, Amer. Meteor. Soc.

7. Tables

Table 1. Number of hours (percentage) in which precipitation was not measured (HPD=0, CRN=0) and the numbers of hours (percentage) when precipitation was measured (HPD>0, CRN>0) in the HPD and CRN networks on days in which no precipitation was recorded at a nearby COOP station.

	HPD = 0	**HPD > 0**	**TOTAL**
CRN = 0	189939 (98.86%)	1619 (0.84%)	191558 (99.70%)
CRN > 0	325 (0.17%)	253 (0.13%)	578 (0.30%)
TOTAL	190264 (99.03%)	1872 (0.97%)	192136 (100%)

8. Figures

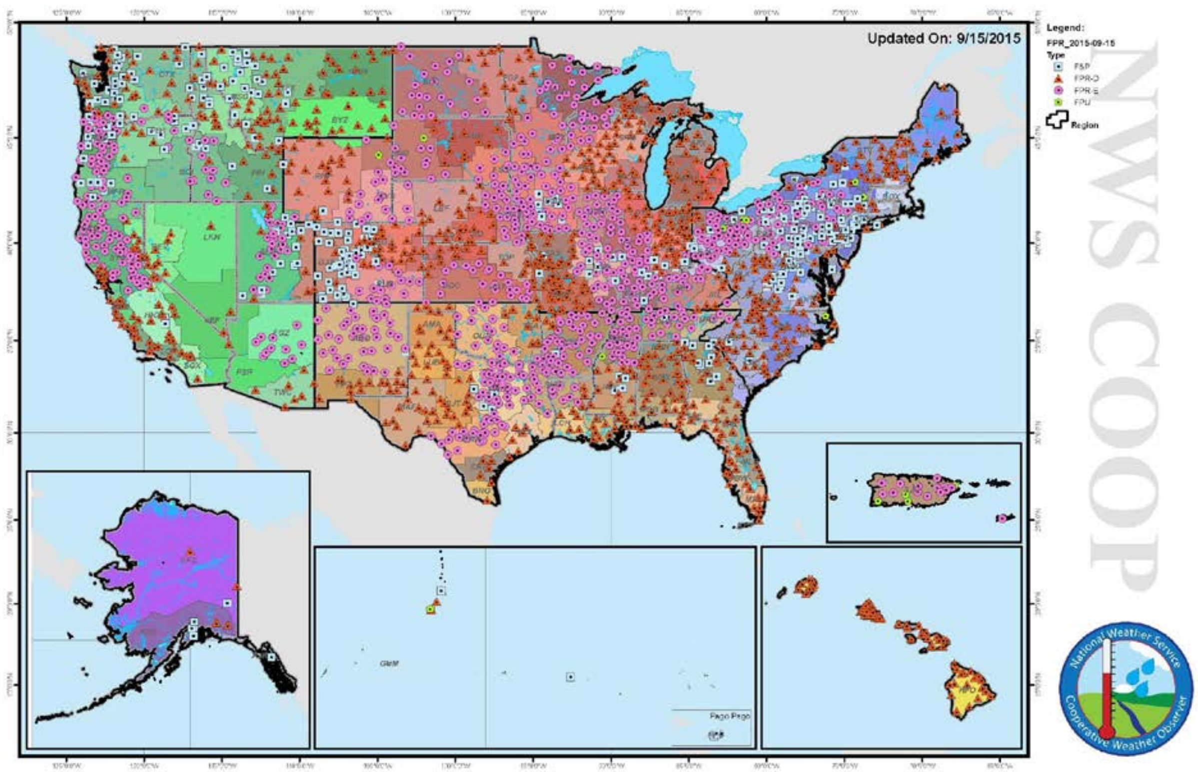

Figure 1. The location of each station in the NWS Fischer-Porter (HPD) network.

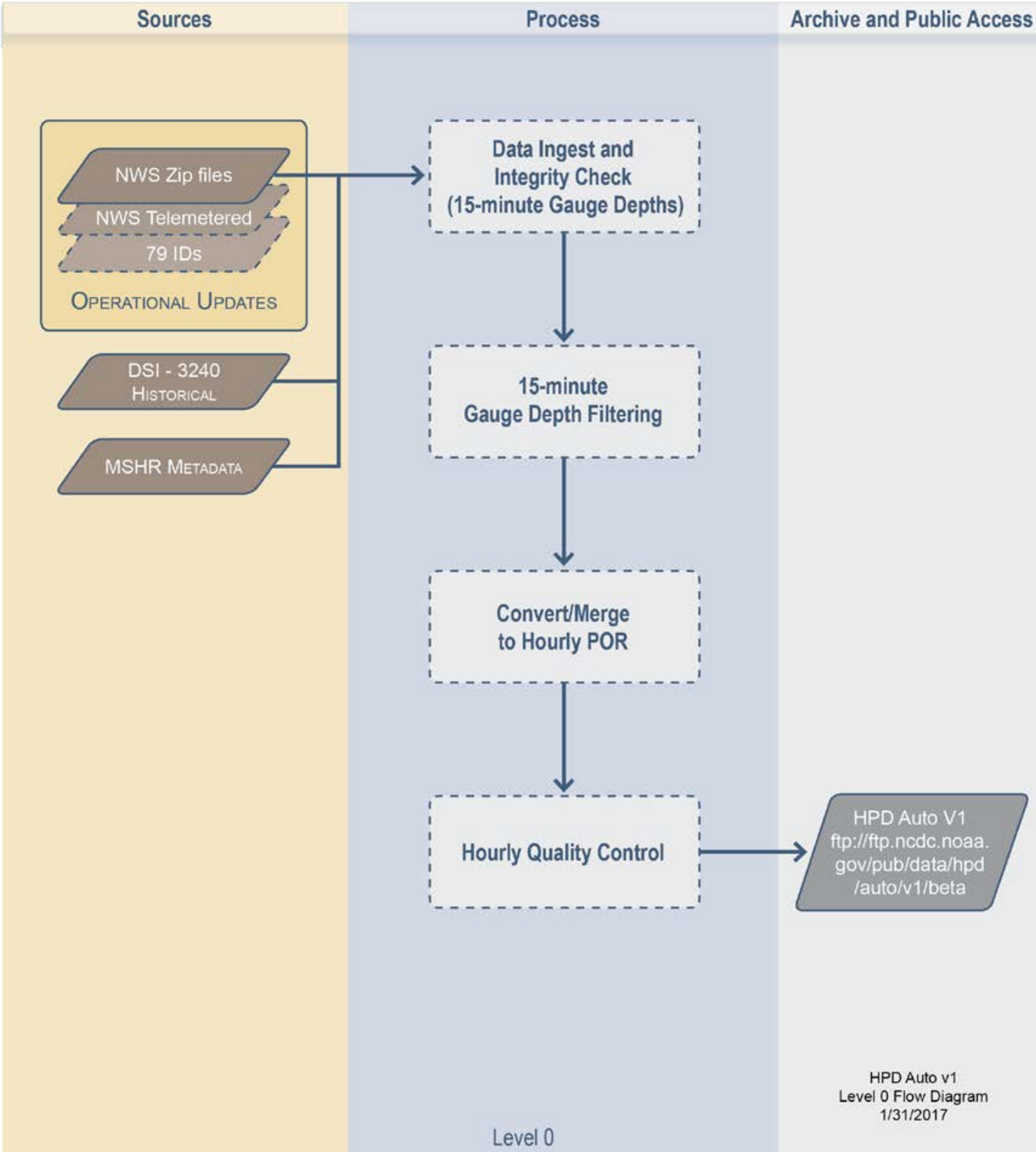

Figure 2. Level 0 flow diagram.

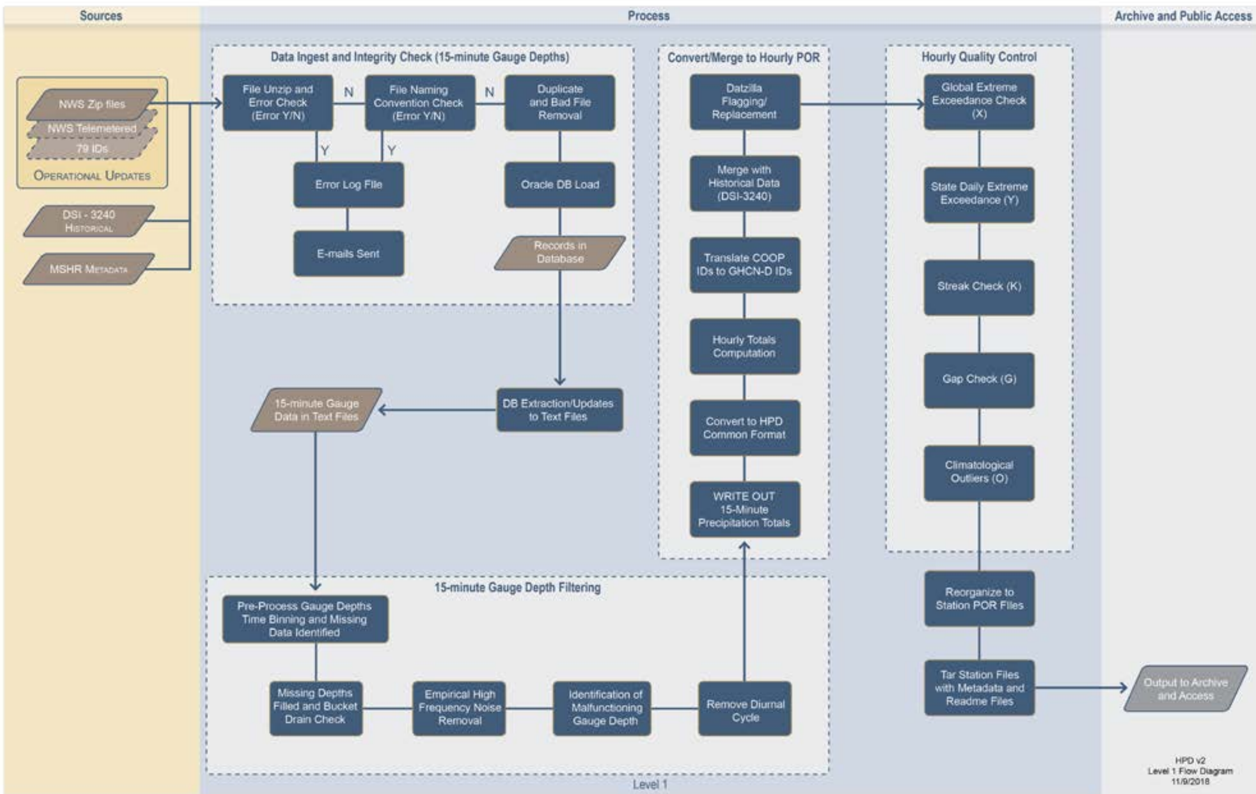

Figure 3. Level 1 flow diagram.

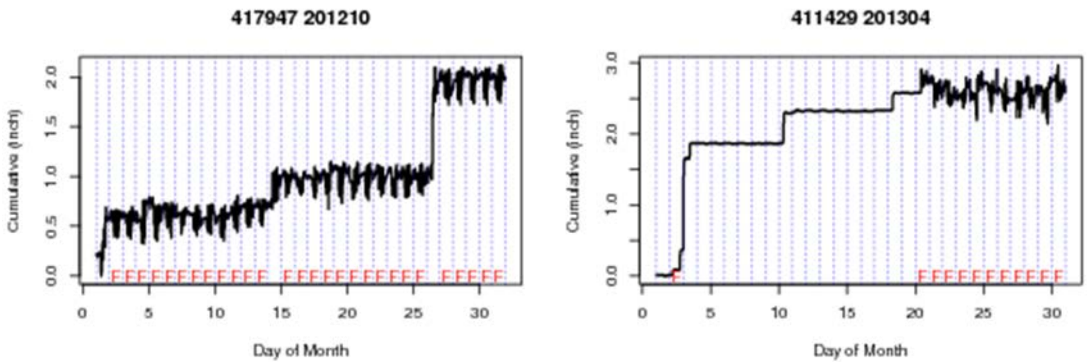

Figure 4. Examples of malfunctioning stations. (a.) F&P gauge depth values for station 417947 during October 2012. Data for most days were invalidated due to high noise levels. Days invalidated are identified with “F”. It was possible to determine the amount of precipitation that fell on 3 days that month (1st, 14th, 26th). (b.) F&P gauge depth values for station 411429 during April 2013. Large fluctuations began suddenly on the 20th day of the month. Days invalidated are identified with “F”. Small diurnal fluctuations are visible on 4-9 and 11-17 April 2013 as discussed in section 3.2.5.

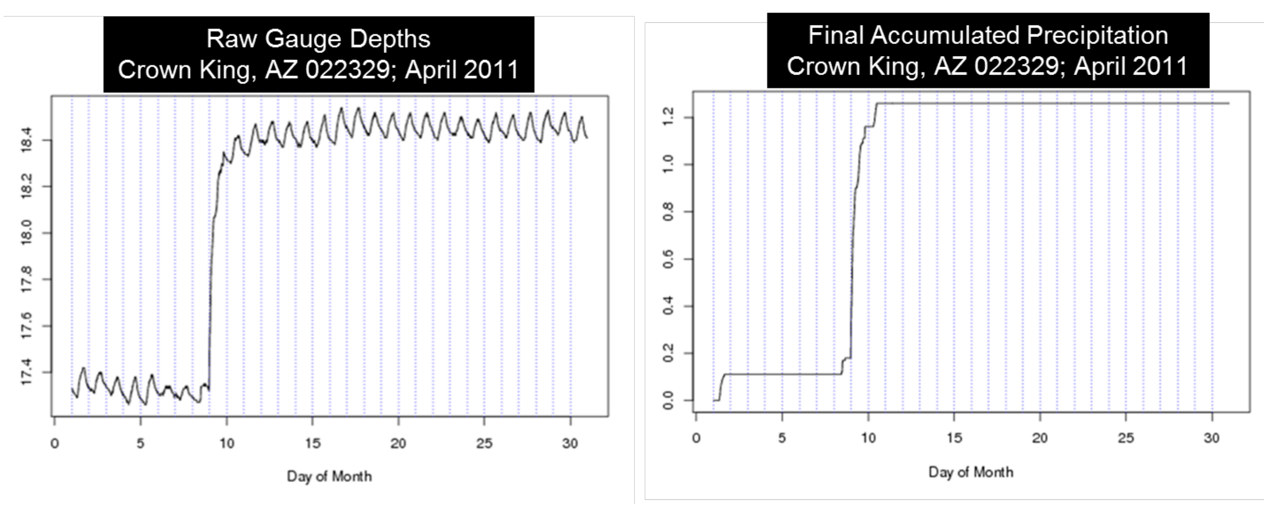

Figure 5. Example of removal of diurnal fluctuations. The diurnal signal is clear in (a) raw gauge depths. As shown in (b), precipitation that fell on 8-9 April is large enough for the amount to be determined. Precipitation on dry days is set to zero.

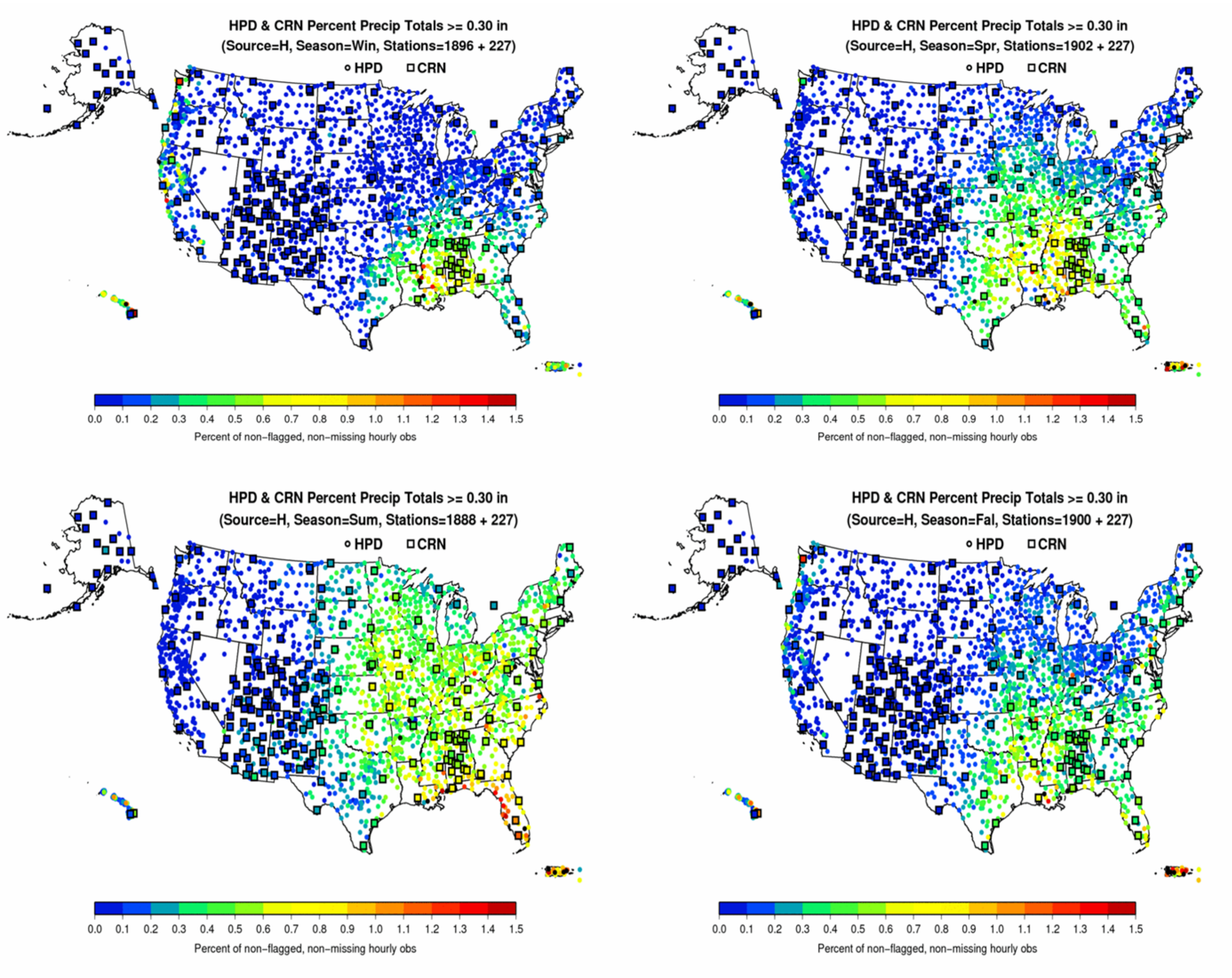

Figure 6. Frequency of occurrence of unflagged/non-missing hourly precipitation >=0.30” (7.6mm) for the HPD and CRN stations; Winter, Spring, Summer, Fall.

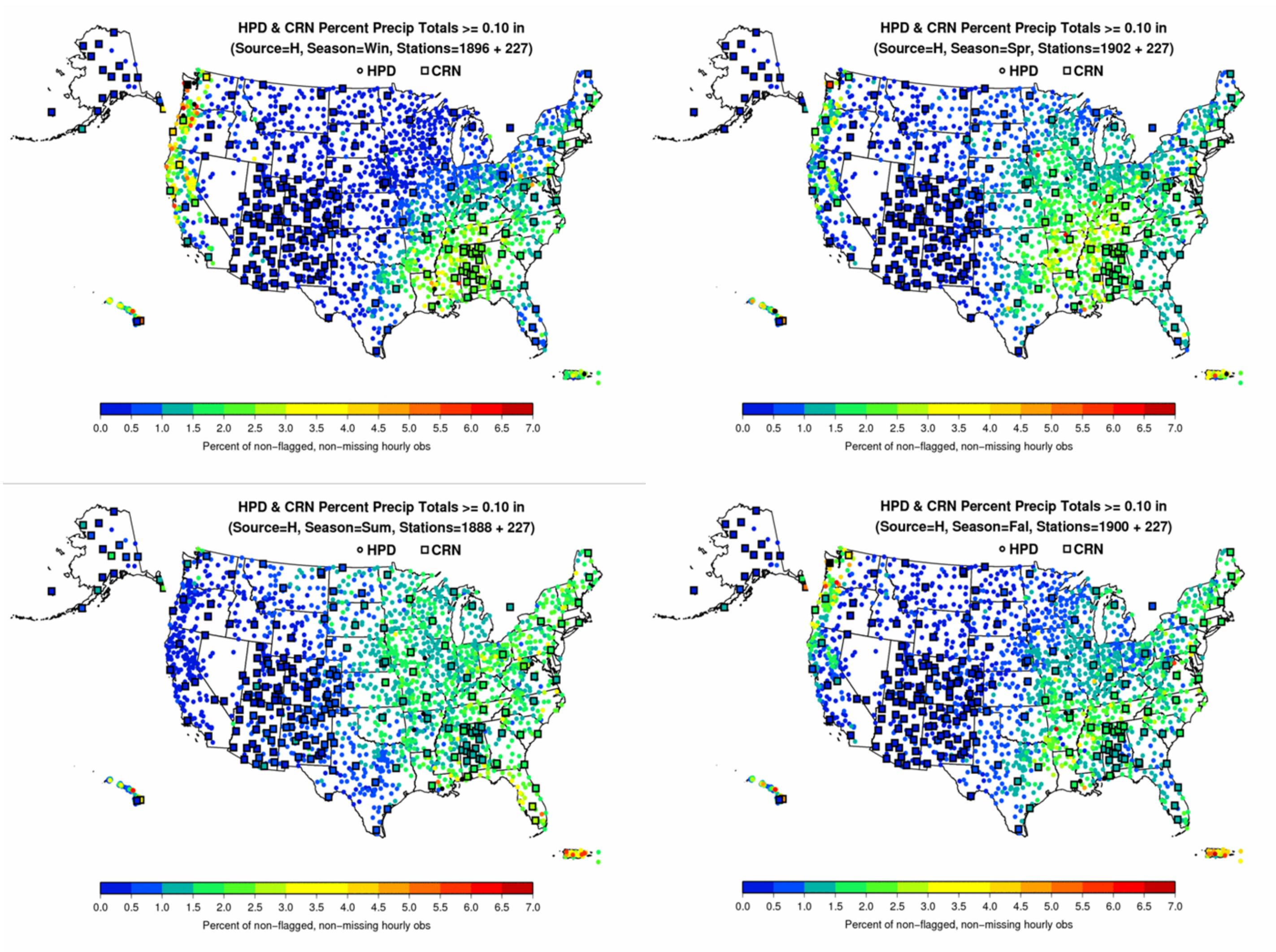

Figure 7. Frequency of occurrence of unflagged/non-missing hourly precipitation >=0.10” (2.5mm) for the HPD and CRN stations; Winter, Spring, Summer, Fall.

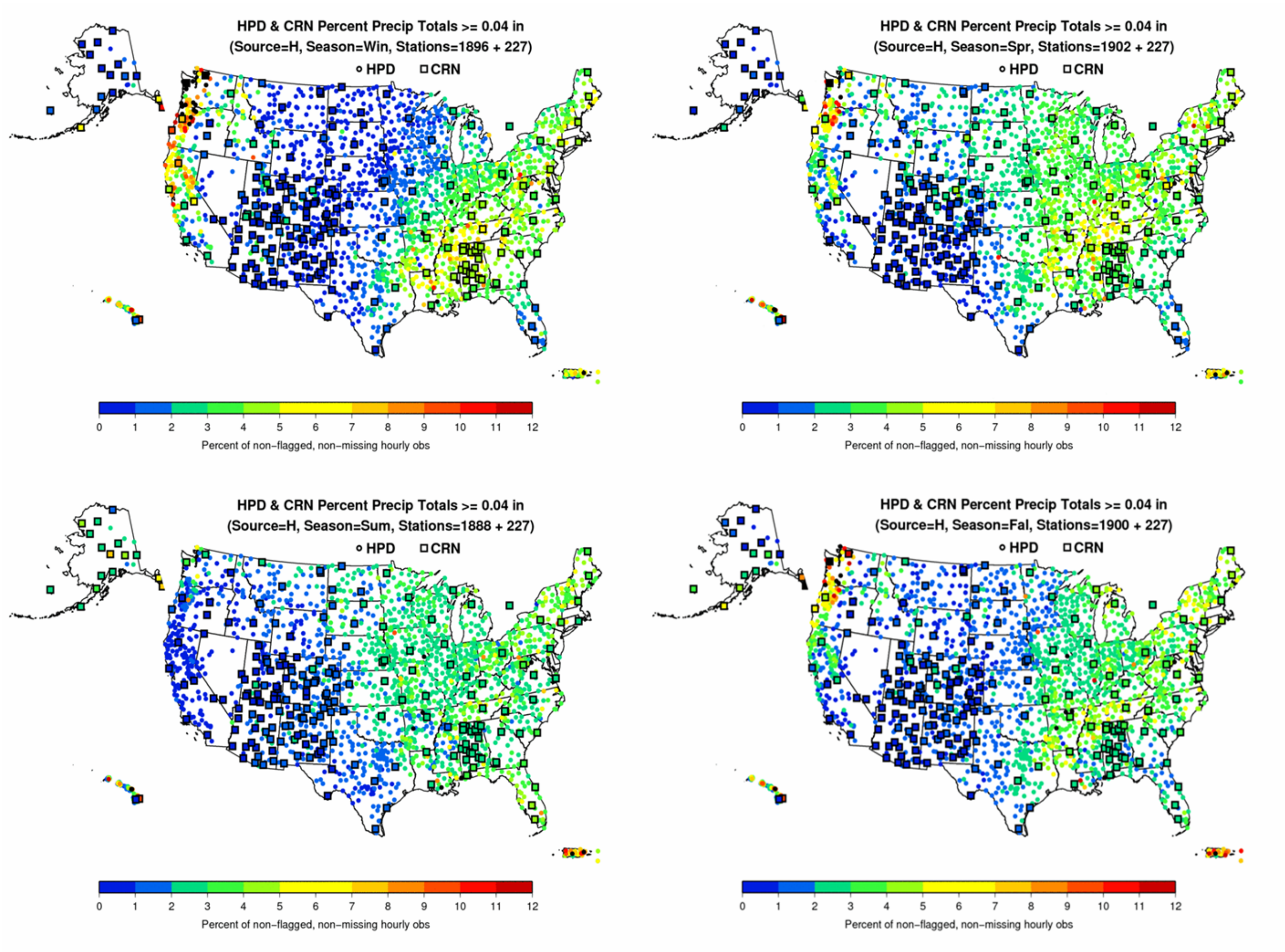

Figure 8. Frequency of occurrence of unflagged/non-missing hourly precipitation >=0.0.4” (1.0mm) for the HPD and CRN stations; Winter, Spring, Summer, Fall.

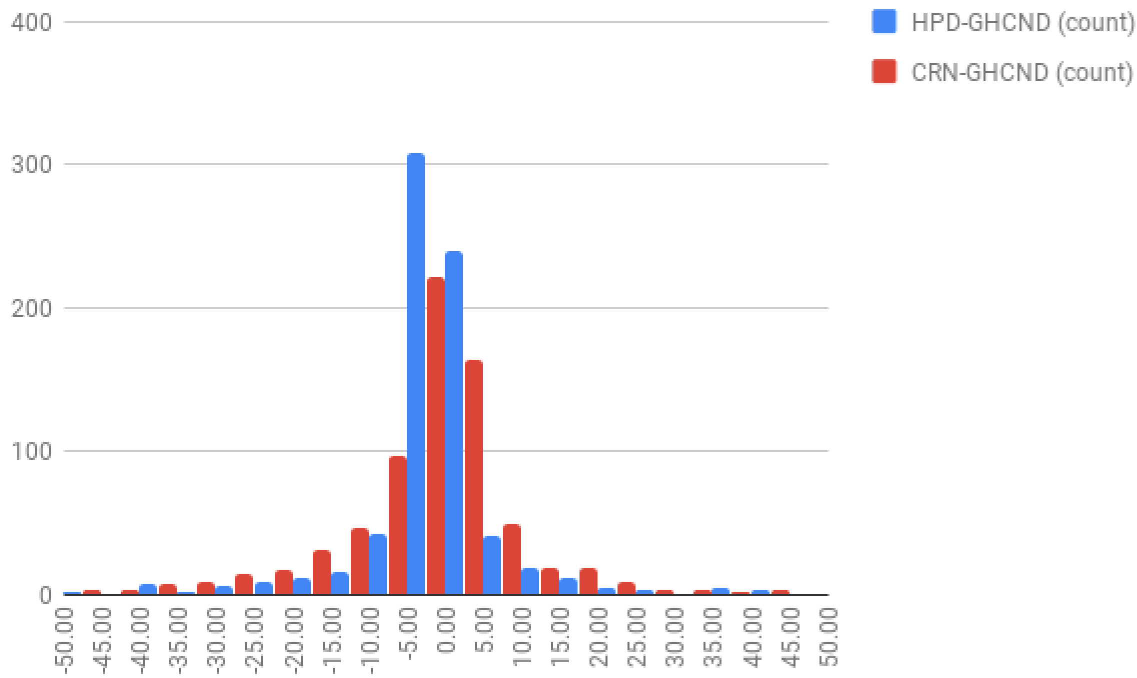

Figure 9. Distribution of differences in daily precipitation totals; HPD minus GHCN-D (blue) and CRN minus GHCN-D (red) for the 743 precipitation events studied.

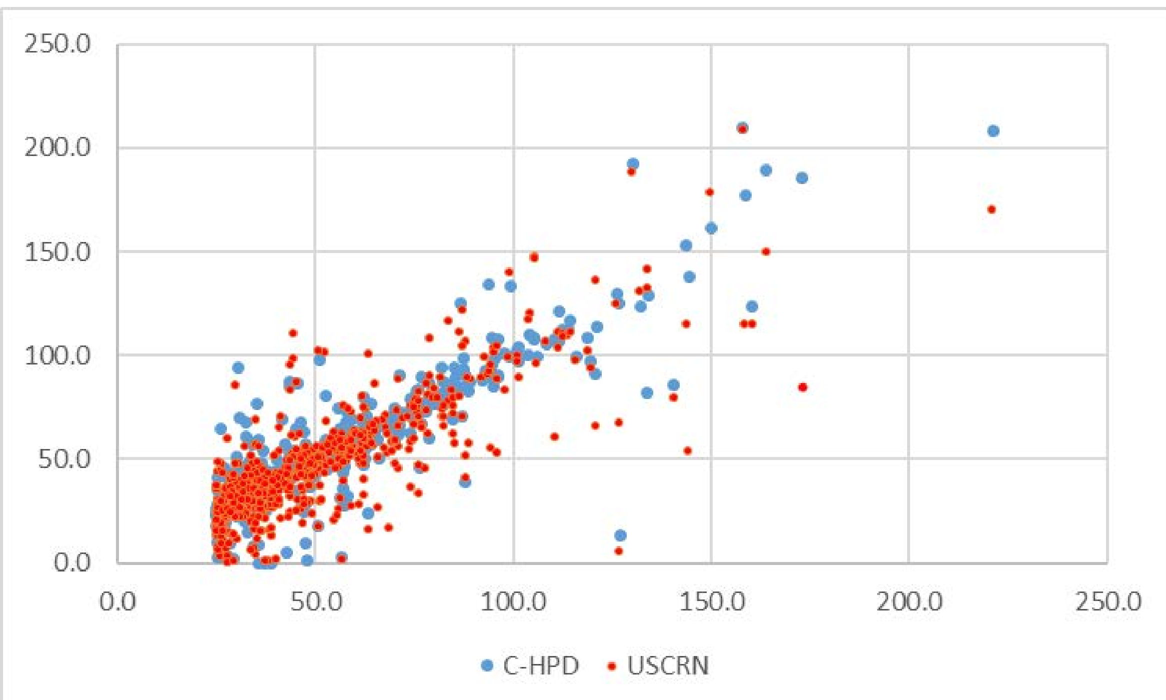

Figure 10. Comparison of GHCN-D event totals against HPD (blue) and CRN (red) for the 743 events studied.