J3.7

OBSERVER BIAS IN DAILY PRECIPITATION MEASUREMENTS AT UNITED STATES COOPERATIVE NETWORK STATIONS

Christopher Daly*, W.P. Gibson, G.H. Taylor, M. Doggett, J.I. Smith
16th AMS Conf. on Applied Climatology, Amer. Meteorological Soc., San Antonio, TX, January 14-18, 2006

1. INTRODUCTION

The Cooperative Observer Program (COOP) was established in the 1890's to make daily meteorological observations across the United States, primarily for agricultural purposes. The COOP network has since become the backbone of temperature and precipitation data that characterize means, trends, and extremes in US climate. COOP data are routinely used in a wide variety of applications, such as agricultural planning, environmental impact statements, road and dam safety regulations, building codes, forensic meteorology, water supply forecasting, weather forecast model initialization, climate mapping, flood hazard assessment, and many others. A subset of COOP stations with relatively complete, long periods of record, and few station moves forms the historical climate network (USHCN), which provides much of the country's official data on climate trends and variability over the past century (Karl et al. 1990, Easterling et al. 1999, Williams et al. 2004).

Precipitation data (rain and melted snow) are recorded manually each day by over 12,000 COOP observers across the US. The measuring equipment is very simple, and has not changed appreciably since the network was established. Precipitation data from most COOP sites are read from a calibrated stick placed into a narrow tube within an 8-inch-diameter rain gauge, much like the oil level is measured in an automobile. The National Weather Service COOP Observing Handbook (NOAA NWS 1989) describes the procedure for measuring precipitation from 8-inch non-recording gauges:

> Remove the funnel and insert the measuring stick into the bottom of the measuring tube, leaving it there for two or three seconds. The water will darken the stick. Remove the stick and read the rainfall amount from the top of the darkened part of the stick. Example: if the stick is darkened to three marks above the 0.80 inch mark (the longer horizontal white line beneath the 0.80), the rainfall is 0.83 inch.

The measuring stick has a large, labeled tick mark every 0.10 inch, a large, unlabeled tick mark every 0.05 inch, and small, unlabeled tick mark every intervening 0.01 inch (Figure 1).

In a recent study, COOP precipitation data were used to extend the work of Johnson et al. (2000) to spatially interpolate input parameters for the GEM6 (Generation of Climate Elements for Multiple Uses) stochastic weather simulation model. Given suitable input parameters, weather simulation models, or "weather generators," can create synthesized daily time series of weather. Spatial interpolation of the input parameters allows daily weather series to be generated at locations where no stations exist. The goal was to expand the original mapping region from a portion of the Pacific Northwest to the entire conterminous US. GEM6 uses a two-state Markov chain of first order for precipitation occurrence, and all other generated quantities are dependent on whether a given day is wet or dry (USDA-ARS 1994). Therefore, it is crucial that the relative frequencies and sequences of wet and dry days be accurately portrayed in the input parameters. In addition, daily precipitation amounts are derived from a mixed exponential distribution that is sensitive to the frequency of observations of precipitation at very low amounts (i.e., less than 1.00 mm).

Initial mapping of some of the precipitation-related weather generator parameters using daily COOP data produced spatial patterns that were highly discontinuous in space, even on flat terrain away from coastlines. When an investigation was conducted to determine the cause of these spatial discrepancies, it was found that precipitation data from most of the COOP stations suffered from serious observer bias, that is, the tendency for the observer to favor or avoid some precipitation values compared to others. Biases included under-reporting of daily precipitation amounts of less than 0.05 inch (1.27 mm), and a strong tendency for observers to favor precipitation amounts divisible by five and/or ten when expressed as inches. These biases were not stationary in time, and thus had significant effects on the temporal trends as well as long-term means of commonly used precipitation statistics. Stations included in the USHCN data set were also affected, raising questions about how precipitation trends and variability from this network should be interpreted.

The objectives of this paper are to make a first attempt at quantifying these biases, provide users of COOP precipitation data with some basic tools and insights for identifying and assessing these biases, and suggest additional investigations and actions to address this issue. COOP observers in the US measure precipitation in English units. Given that the observer bias discussed here is uniquely tied to this system, precipitation amounts are given first in inches, followed

* *Corresponding author address*: Christopher Daly, PRISM Group, Oregon State University, 326 Strand Ag Hall, Corvallis, OR 97331; email: daly@coas.oregonstate.edu

by millimeter equivalents in parenthesis. All other measures are given in standard metric, or MKS, units.

2. TYPES OF OBSERVER BIAS AND ASSESSMENT STATISTICS

Two major types of observer bias were discovered in this initial investigation: (1) So-called “under-reporting bias,” or under-reporting of daily precipitation amounts of less than 0.05 inch (1.27 mm); and (2) so-called “5/10 bias,” or over-reporting of daily precipitation amounts divisible by five and/or ten, such as 0.05, 0.10, 0.15, 0.20, and 0.25 inch (1.27, 2.54, 3.81, 5.08, and 6.35 mm). These two types are usually related; a station with a 5/10 bias is likely to have an under-reporting bias, as well.

The source of the daily precipitation data analyzed in this study was the National Climatic Data Center’s (NCDC) TD3200 data set (NOAA-NCDC 2006). Each station was subjected to data completeness tests of sufficient rigor to ensure reasonable GEM6 parameters, given good-quality data. To ensure the accurate calculation of wet/dry day probabilities, daily precipitation entries that were flagged as accumulated totals for more than one day were set to missing. For a given year to be complete, each of the 26 14-day periods in the year had to have at least 12 days (~85%) with non-missing data, and there had to have been at least 26 (~85%) complete years within the 1971-2000 period. GEM6 operates on 14-day statistical periods, hence the use of this time block. Tests were conducted using 90% thresholds for data completeness, but the number of stations passing such stringent tests were very low, and data quality did not improve noticably.

Two kinds of simple statistical tests were devised to detect stations that exhibited one or both observational biases. The under-reporting bias test consisted of calculating the ratio:

$$R_L = C_{6\text{-}10} / C_{1\text{-}5} \qquad (1)$$

where $C_{6\text{-}10}$ is the observation count in the 0.06 - 0.10-inch (1.52 – 2.54-mm) range, $C_{1\text{-}5}$ is the observation count in the 0.01-0.05 inch range, and R_L is the ratio of the two. A station exhibiting an R_L that exceeded a given threshold was most likely under-reporting precipitation in the 0.01 – 0.05-inch (0.25 – 1.27-mm) range.

A station’s 5/10 bias was assessed by separating the frequencies of observations in amounts divisible by five and/or ten hundredths of an inch and those not divisible by five or ten hundredths of an inch into separate populations, and comparing their means. If they were significantly different, a 5/10 bias was indicated. In order to make consistent comparisons across a spectrum of frequency bins, it was necessary to de-trend the frequency histogram. This was done by fitting each station’s precipitation frequency histogram to a gamma distribution (Evans et al. 2000). Given that the distribution of precipitation is highly variable for different parts of the US, rather than fitting a single distribution to all stations, a gamma distribution was fit to each station. It was not necessary that the gamma function fit the data precisely, or without bias; rather, the predictions were used only as a way to de-trend the frequency distribution.

Due to computational constraints in solving the gamma distribution, predictions become unstable as precipitation approaches zero. Therefore, no frequency predictions were made below 0.03 inch (0.76 mm). (This lower bound has no effect on the detection of under-reporting bias, because frequencies were de-trended for the 5/10 test only.) In addition, no frequency predictions were made above one inch (25.40 mm), because observed frequencies at these precipitation amounts were typically very low.

A station’s 5/10 bias was assessed by separating the frequencies of observations in amounts divisible by five and/or ten hundredths of an inch and those not divisible by five or ten hundredths of an inch into separate populations, and comparing their means. If they were significantly different, a 5/10 bias was indicated. In order to make consistent comparisons across a spectrum of frequency bins, it was necessary to de-trend the frequency histogram. This was done by fitting each station’s precipitation frequency histogram to a gamma distribution (Evans et al. 2000). Given that the distribution of precipitation is highly variable for different parts of the US, rather than fitting a single distribution to all stations, a gamma distribution was fit to each station. It was not necessary that the gamma function fit the data precisely, or without bias; rather, the predictions were used only as a way to de-trend the frequency distribution.

Due to computational constraints in solving the gamma distribution, predictions become unstable as precipitation approaches zero. Therefore, no frequency predictions were made below 0.03 inch (0.76 mm). In addition, no frequency predictions were made above one inch (25.4 mm), because observed frequencies at these precipitation amounts were typically very low.

The percent difference, or residual (R), between expected and observed frequencies was calculated as follows:

$$R = 100 * (P - O) \qquad (2)$$

where P is the predicted frequency (via the gamma function) and O is the observed frequency. Fives and tens bias were tested separately. For the fives bias test, the first residual mean was calculated by averaging the residuals over the so-called “ones bins,” which include all amounts, except those divisible by five; the second was calculated as the average of all residuals for the so-called “fives bins,” which include only amounts divisible by five. The two means were computed similarly for the tens bias test. A standard comparison of the means for small samples was then applied, and t-statistic calculated. The residuals from the predicated and observed frequencies were reasonably normally distributed, even for severely biased histograms.

Although our main interest was in cases for which the fives-bin residuals were significantly greater than the ones-bin residuals, a two-tailed t-test was applied in case there were situations for which the opposite was true, suggesting an avoidance of the fives bins. This occurred only rarely.

The under-reporting bias and 5/10 tests were run on COOP stations that passed the data completeness tests discussed earlier. Initial threshold alpha values for the 5/10 bias tests and ratio cutoffs for the under-reporting bias test were set, and stations which failed any of the tests were removed from the data set. The station values were then examined spatially in parts of the country where terrain and coastal features should have minimal effect on the spatial patterns of precipitation. The process of setting threshold values, removing stations, and mapping the remaining stations was performed repeatedly until it appeared that an optimal balance between removing the worst stations and keeping the best stations had been reached. The final threshold alpha level for the 5/10 t-tests was 0.01, and the final threshold ratio for the under-reporting bias test was 0.6. These threshold values are assumed throughout this paper.

3. EXAMPLES OF OBSERVER BIAS

Figure 2 depicts frequency histograms of daily precipitation amounts from two COOP stations that show no visible observer bias during the period 1971-2000, and passed all observer bias tests (Table 1). Bishop, CA (COOP ID 040822), a desert site with a mean annual precipitation of 4.90 inches (125 mm), and Quillayute, WA (456858), a coastal rain forest site with a mean annual precipitation of 101.90 inches (2588 mm). Despite representing extremely different precipitation regimes, the histograms have a remarkably similar shape. Both stations exhibit a maximum frequency at 0.01 inches (0.25 mm), with a relatively smooth decrease in frequency of occurrence as the daily precipitation amount increases. The Bishop station experienced fewer precipitation events than the Quillayute station, and thus it is not surprising that the Bishop histogram is not as smooth as the Quillayute histogram. Generally, the more precipitation events included, the smoother the appearance of the frequency histogram; at least ten years of data are typically required to obtain a smooth histogram at an unbiased site, and to provide enough frequency counts in various precipitation bins to produce stable statistical results.

Figures 3-5 show examples of frequency histograms from seriously biased stations paired with those from nearby stations with little bias. Accompanying observer bias test results are given in Table 1. All stations discussed passed the data completeness tests for the 1971-2000 period. These comparisons, and others analyzed but not shown here, strongly suggest that the unusual frequency histograms are not a result of true climatic conditions, but of inaccurate reporting of those conditions. Philadelphia, MS (COOP ID 226894; Figure 3a) suffers from a considerable under-reporting bias (Table 1). Instead of the expected decreasing frequency trend between 0.01 inch (0.25 mm) and 0.10 inch (2.54 mm), Philadelphia exhibits a sharply increasing frequency trend, with absolutely no observations of 0.01 inch (0.25 mm) during the thirty-year period. This station also suffers from 5/10 bias (Table 1), with several frequency spikes at amounts divisible by ten. The number of observations of 0.10 inch (2.54 mm) is strikingly high. The observer also seemed to have avoided readings ending in nine, such as 0.29, 0.39, and 0.49 inches (7.40, 9.90, and 12.40 mm, respectively), possibly rounding up to the nearest 0.10 inch (2.54 mm). In contrast, the frequency histogram at Laurel, MS (224939), approximately 100 km to the south, shows only a slightly visible 5/10 bias (Figure 3b) and passes all observer bias tests (Table 1).

Figure 4 compares precipitation frequency histograms from Purcell 5 SW, OK (347327) and (b) Watonga, OK (439364), approximately 150 km to the northwest. Purcell 5 SW has a considerable under-reporting bias, as well as a well-defined 5/10 bias, and fails all observer bias tests (Table 1). In contrast, Watonga shows little under-reporting bias and perhaps only a slight 5/10 bias, and passes all observer bias tests (Table 1). The frequency of daily precipitation values of 0.01-0.03 inch (0.25-0.762 mm) was two to three times lower at Purcell 5 SW than at Watonga, and this pattern continued into the trace category (not shown). Further analysis showed that the frequency deficit was made up by a relative increase in the percent of days with zero precipitation. Despite Purcell 5 SW receiving about twenty-five percent more precipitation annually than Watonga, both stations recorded zero precipitation on about the same number of days. This suggests that the observer at Purcell 5 SW had a higher threshold for “inconsequential” precipitation than the observer at Watonga (Hyers and Zintambila 1993, Snijders 1986).

Figure 5 compares Cloverdale, OR (351682), on the northern Oregon coast, with Otis 2 NE, OR (356366), 20 km to the south, and Astoria Airport, OR (350328), 100 km to the north. Cloverdale (Figure 5a) exhibits a striking 5/10 bias, with readings divisible by five and ten occurring three to six times more often than other amounts. This station failed all observer bias tests (Table 1). The frequency histogram for Otis 2 NE (Figure 5b) is remarkably different, with only a minor 5/10 bias visible. Interestingly, Otis failed the fives bias means test by a small margin (Table 1), indicating that the observer bias tests at the current 0.01 alpha threshold identified stations with biases that were not prominent visually. Astoria’s frequency histogram

(Figure 5c) shows no appreciable bias and passed all tests (Table 1). Precipitation at Astoria Airport was observed by trained personnel at the Weather Service Forecast Office until the measurement system was converted to ASOS (Automatic Surface Observing System) in 1993. The conversion to ASOS is noteworthy, because at that time, precipitation began to be measured electronically, minimizing the potential for human observing bias.

4. TEMPORAL RAMIFICATIONS

The severity of observer bias was often found to vary over time. Consider USHCN station Vale, OR (358797) and Malheur Branch Experiment Station (355160; hereafter abbreviated as Malheur), located approximately 20 km apart in eastern Oregon. For the 1971-2000 period, Vale exhibits a subtle low bias and a significant 5/10 bias, while Malheur exhibits only a small 5/10 bias (not shown). Vale failed all three observer bias tests, while Malheur passed all three (Table 1). An example of temporal variability in observer bias at Vale is shown in Figure 6. The period 1930 to 1950 exhibited visible under-reporting and 5/10 biases (Figure 6a), 1950-1980 was relatively free of bias (Figure 6b), and 1980-2005 returned to an under-reporting bias and a 5/10 bias (Figure 6c).

Such changes over time complicate the issue by presenting a "moving target" to efforts to assess and adjust for observer bias. However, they also provide valuable insight into the implications of observer bias by allowing analysis of the relationships between trends in the observer bias test statistics and trends in commonly-used precipitation statistics at two nearby stations. Figure 7 shows time series of the 10-year running mean of R_L, the under-reporting bias ratio, and the maximum of the fives and tens bias test t-statistics (t_{510}), for Vale and Malheur for the period 1965-2004. A 10-year running mean is used, because at least 10 years of data are typically required for stable statistical results. The t_{510} statistic reflects the highest t-statistic (worst case) of the two 5/10 tests, and a ten-year running mean is used to capture a sufficient number of observations to produce a statistically stable test result. The period 1965-2004 was chosen because it was characterized by a rapid divergence in the trends of R_L and t_{510} at the two stations. Vale showed a clear trend towards increasing observer bias in both test statistics, while Malheur remained reasonably unbiased throughout the period.

Figure 8 presents time series trends for three commonly-used precipitation statistics at Vale and Malheur: the percent of days that were wet, average precipitation on a wet day, and the mean annual precipitation. The percent of wet days at both stations began at about 19 percent during the 10-year period ending in 1974, but diverged sharply in later years, with Vale trending strongly downward and Malheur slightly upward (Figure 8a). By the 10-year period ending in 2004, the percent of wet days at Vale had reached a low of 13 percent, a 6-percent drop, while Malheur reported an increase to about 25 percent, a 7-percent rise. Trends in the average precipitation on a wet day show a near-doubling of the average daily precipitation at Vale from 0.12 to 0.22 inches (3.05 mm to 5.59 mm) per day, while Malheur shows a slight drop (Figures 8b). Trends in mean annual precipitation, a relatively stable precipitation statistic, did not exhibit dramatically different trends, but Vale's value increased relative to Malheur (Figure 8c).

Relationships between the temporal trends in R_L and t_{510} and those of the three precipitation statistics in Figure 8 were explored by generating scatter plots of the inter-station differences of one versus the other. As seen in Figure 9, definite relationships exist. As the difference in R_L increased, signaling increased under-reporting bias at Vale, there was a strong linear tendency for the number of wet days to decrease compared to Malheur (Figure 9a). This suggests that the observer increasingly recorded precipitation values of zero or trace on many days, up to 13 percent more than recorded at Malheur by the 10-year period ending in 2004. Further analysis revealed that the number of zero precipitation days was strongly and positively related to trends in R_L, while the number of trace days was not, suggesting that the observer recorded more zeros than actually occurred.

Given the decreasing wet-day trend at Vale, it was not surprising to see an increase in the average precipitation on a wet day (Figure 9b). This statistic represents the average precipitation intensity when precipitation occurs. If there are fewer wet days and the total amount of annual precipitation remains reasonably constant, it stands to reason that the precipitation intensity would rise. However, observer bias did affect the mean annual precipitation, as well (Figure 9c). The relationship was strongest with changes in t_{510}; for every increase in t_{510} difference of 1.0, the mean annual precipitation increased by 0.16 inch (4 mm), up to about 1 inch (25.4 mm). Given that the mean annual precipitation at these two stations was approximately ten inches, the increase in mean annual precipitation at Vale attributable to observer bias was about ten percent. Reasons for this relationship are not clear, but may have been related to the observer rounding to higher values divisible by five and ten.

5. SCOPE OF OBSERVER BIAS

To gain perspective on the extent of observer bias in COOP precipitation data across the continental US, all COOP stations having at least some data within the 1971-2000 period were subjected to the data completeness and observer bias test tests. The results, summarized in Table 2 and Figure 10, were not encouraging. Out of over 12,000 candidate COOP stations, 25 percent passed the data completeness tests, and of those, 25 percent passed the observer bias tests, leaving just over six percent of the total, or 784 stations (Table 2). USHCN stations, included in this network partly because of their long and complete records, faired better on the data completeness tests, with two thirds passing. However, the observer bias failure rate was about the same as for the total COOP population. In the end, 18 percent, or 221 USHCN stations, passed all tests (Table 2). The spatial distribution of USHCN stations passing and failing the tests showed no particular pattern across the continental US, with all major regions and climate regimes affected (Figure 9).

As a check on the reasonableness of the observer bias screening tests, daily precipitation totals from hourly data at 224 first-order stations were obtained from Surface Airways Observation archives for the period July 1, 1996 – July 31, 2006. To provide sufficient observations for testing, only stations with at least one hundred wet days and a total of 3000 non-missing days were accepted. Given that all of these stations employed the ASOS automated observing system during this period, they should not suffer from observer bias. All of these stations passed the observer bias tests, suggesting that these tests, while designed as rough screening devices, were providing reasonable results.

6. DISCUSSION AND RECOMMENDATIONS

The causes of observer bias are not yet clear, but some early speculations can be made. One cause of the 5/10 bias may be the way that the measuring sticks are marked and labeled; the larger the mark or label at a given amount, the more likely an observer will choose that amount. Another possible contributing factor is that not all COOP measuring sticks are alike. At the Corvallis Hyslop COOP station, the observer possesses two measuring sticks, an old one issued by the "U.S. Weather Bureau,", and a newer one issued by the "National Weather Service." He uses the older stick almost exclusively, because the finish on the new one is too smooth and impervious to allow water to impregnate the stick sufficiently to create a darkened, wetted area that is easy to read. The water beads up and runs off the front surface of the newer stick, forcing the observer to turn the stick to the side and read one of the narrow edges, where the water soaks in a bit further. Given that the larger and labeled tick marks on the stick represent "round" values that an observer might already gravitate towards, this additional measuring uncertainty may further motivate observers to choose one of these round numbers.

There also appears to be a strong tendency for observers to choose 0.1 inch (2.54 mm) at the expense of lower values, with the occasional exception of 0.05 inch (1.27 mm). It is possible that many observers may not see the need to take a precipitation measurement if they perceive that "inconsequential" precipitation had fallen in the last 24 hours. They may record zero for such days, and allow what is effectively an accumulation to occur until an observation is made. This is consistent with our analysis showing a disproportionately large number of zero observations at highly biased stations. It appears that on many occasions, the lower limit of "consequential" precipitation is in the vicinity of 0.10 inches (2.54 mm), which may be rounded up to that value if the observer read the measuring stick with a 5/10 bias.

Regardless of the exact reasons, observer bias suggests a lack of understanding of COOP precipitation measurement procedures, a lack of commitment to faithfully carry them out, or both. These problems may be largely unavoidable given the volunteer status of the COOP observers, and the lack of accountability associated with this status. However, if COOP data are to be used in what has become an increasingly large, diverse, and critical set of applications, there is a correspondingly heightened need to improve the quality of these data. One possible step would be to develop training materials that put procedural instruction in the context of data applications. Do observers know how their data are being used? Do they know that recording a zero on days with just a little rainfall can compromise the results of applications that use their data? Do they understand the meaning of such terms such as precision and accuracy and why they are important in real-world applications?

Unfortunately, even the most effective training materials are not likely to eliminate observer bias. One solution to human observer bias might be to automate the COOP precipitation measurement system. NOAA's Environmental Real-time Observation Network (NERON) is a new national program designed to accomplish this. In the first phase of NERON, one hundred automated stations were installed in New England and eastern New York. As stated in NERON documentation, the main advantage of automating the COOP network is the availability of air temperature and precipitation every five minutes and disseminated in real time (NOAA-NWS 2006). However, automation is expensive, and could introduce other sources of bias inherent in automated instrumentation, such as electronic biases and instrument malfunctions, and potential difficulty with frozen precipitation and heavy precipitation events.

Observer bias is not easily identified by quality control procedures running on a day-by-day, or even month-by-month, basis. Our initial analysis suggests that observer bias is not characterized by extremely high measurements on low precipitation days, or by very low precipitation measurements on high precipitation days. Instead, the biased values are "in the ball park," and differences between neighboring stations are typically swamped by the spatial variability of precipitation, and the complicating factor of variable times of observation among nearby stations. The effects of observer bias accumulate over time, and unless the bias is extremely obvious, only become visible through analysis of long-term statistics.

While the effects of observer bias are most easily identified with long-term statistics, the phenomenon itself is temporally complex and unpredictable. Many COOP stations engage two or more observers who may be responsible for recording data on different days, or fill in for one another during travel and vacation times, in addition to periodic turnover of the personnel themselves. This can lead to a confusing spectrum of biases with complex temporal behaviors at the same station. Observer changes are typically not noted in the standard COOP metadata, and thus cannot be easily tracked over time. However, even at stations operated by a single observer over many decades, our analyses have shown that distinct and significant temporal trends in observer bias can still occur.

This study has only scratched the surface of this issue, and it will take much more work to adequately assess the true scope and implications of observer bias on a variety of precipitation statistics. Additional studies should seek to better characterize the nature of observer bias, develop more robust statistical tests to identify various types of observer bias, and possibly develop an early warning system to identify stations that are beginning to show increases in observer bias. At the least, confidence intervals around the means and temporal trends in precipitation statistics calculated from these stations need to be estimated. One possible approach is to use data from longer-term automatic observing systems to gain more insight into the implications of observer bias for various precipitation statistics, and how they might be accounted for. Candidate systems are ASOS, and high-quality, smaller-scale automated networks that have been running for at least ten years (e.g., the Oklahoma Mesonet), to allow the calculation of stable, long-term precipitation statistics. However, the biases inherent in automated systems discussed previously would have to be accounted for.

The authors have developed an observer bias web application that allows users to create a frequency histogram of daily precipitation observations at any COOP station over any time period for which the PRISM Group data base has information. This application can be accessed by the public at http://mistral.oce.orst.edu/bias/.

7. REFERENCES

Easterling, D. R., T. R. Karl, J. H. Lawrimore, and S. A. Del Greco, 1999: United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data for 1871-1997. ORNL/CDIAC-118, NDP-070. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee. 84 pp.

Evans, M., N.A.J. Hastings, J. B. Peacock, 2000: *Statistical Distributions*. Wiley, 98-105.

Hyers, A.D., and H.J. Zintambila, 1993: The reliability of rainfall data from a volunteer observer network in the central United States. *GeoJournal*, 30.4, 389-395.

Johnson, G.L., C. Daly, C.L. Hanson, Y.Y. Lu and G.H. Taylor. 2000: Spatial variability and interpolation of stochastic weather simulation model parameters. *Journal of Applied Meteorology* 39, 778-796.

Karl, T.R., C.N. Williams, Jr., F.T. Quinlan, and T.A. Boden, 1990: United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data, Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

NOAA-NCDC, 2006: Daily surface data, DS3200. National Climatic Data Center, National Oceanic and Atmospheric Administration. Online data access. http://cdo.ncdc.noaa.gov/pls/plclimprod/poemain.accessrouter?datasetabbv=SOD

NOAA-NWS, 1989: *Cooperative Station Observations*. National Weather Service Observing Handbook No. 2, Observing Systems Branch, Office of Systems Operations, Silver Spring, MD, July 1989, page 19. http://www.nws.noaa.gov/om/coop/Publications/coophandbook2.pdf

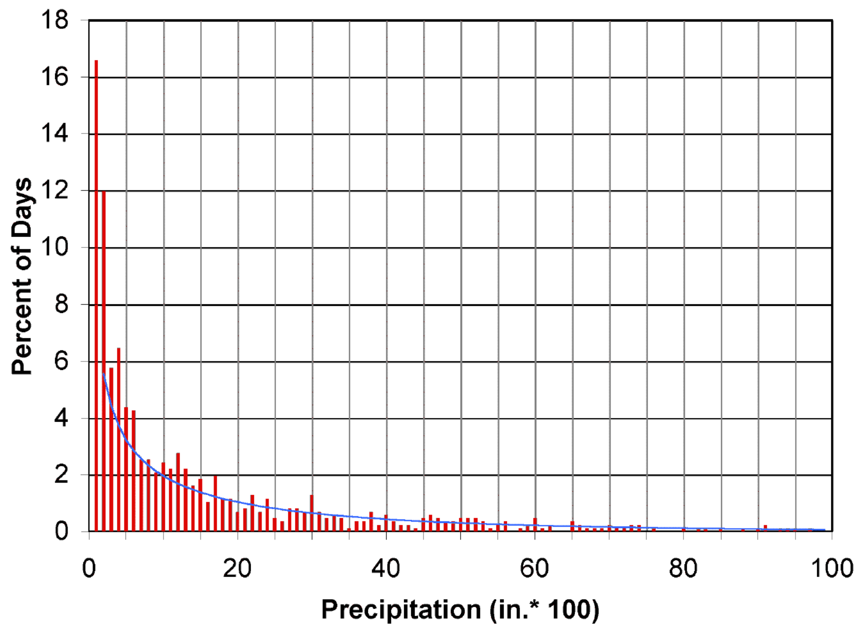
NOAA-NWS, 2006. Building NOAA's Environmental Real-time Observation Network. Draft guidelines for NERON program. Silver Spring, MD. http://www.isos.noaa.gov/documents/Files/NERON PFRD Draft 0.6.pdf

Snijders, T.A.B., 1986: Interstation correlations and nonstationarity of Burkina Faso rainfall. *Journal of Climate and Applied Meteorology*, 25, 524-531.

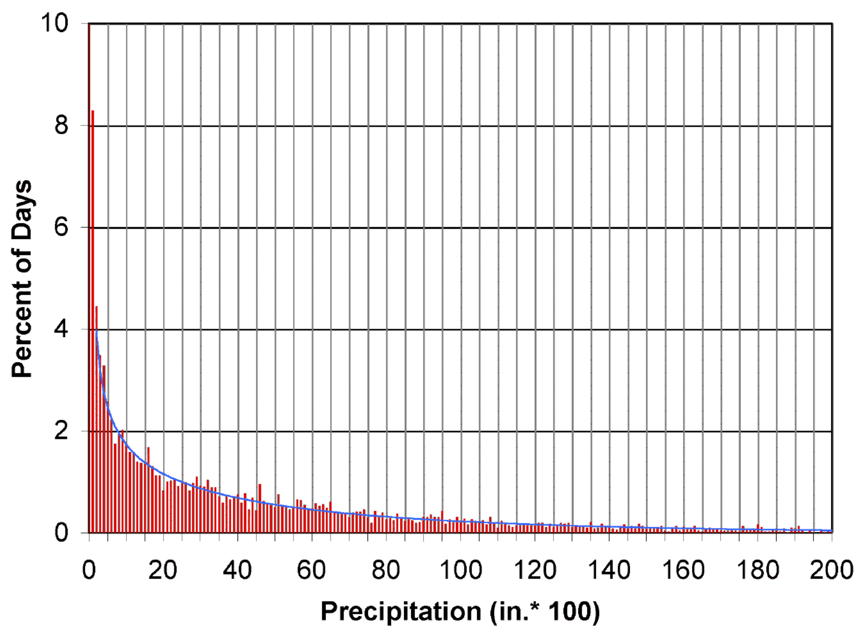
USDA-ARS, 1994: Microcomputer Program for Daily Weather Simulation in the Contiguous United States. *ARS-114.* July 1994. United States Department of Agriculture, Agricultural Research Service.

Williams, C. N., R. S. Vose, D. R. Easterling, and M. J. Menne, 2004: United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data. ORNL/CDIAC-118, NDP-070. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Figure 1. Standard measuring stick used to record precipitation in a COOP rain gauge.

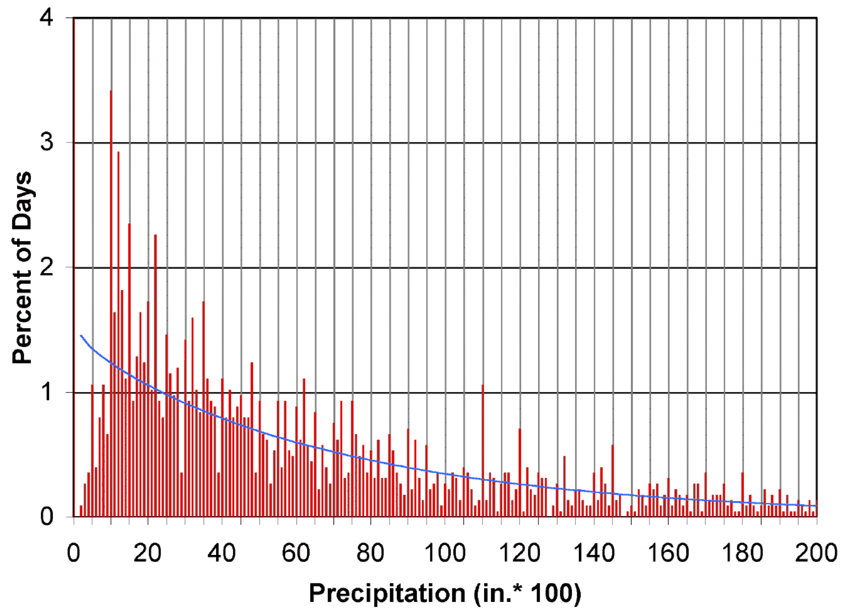


(a)

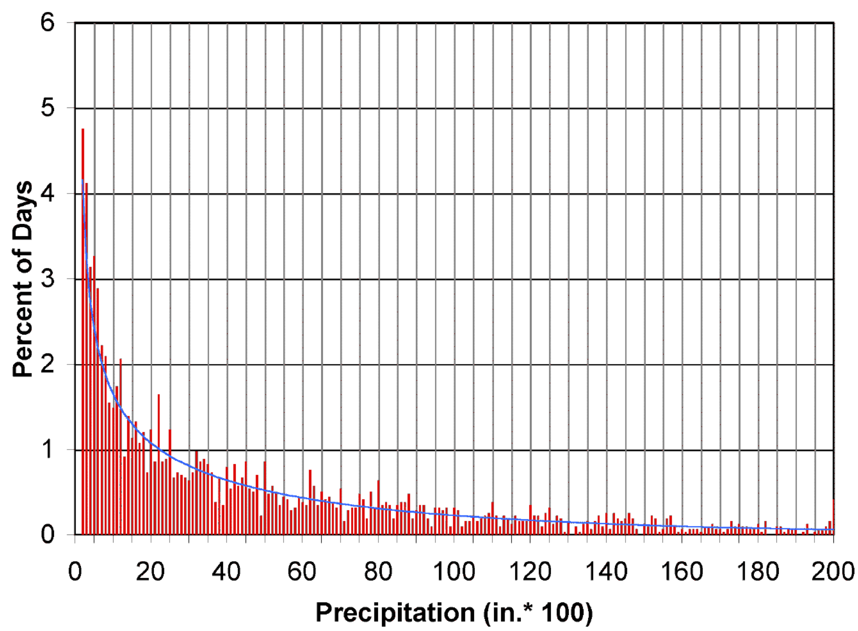


(b)

Figure 2. Percent frequency distribution of daily precipitation of a least 0.01 inch for the period 1971-2000 at COOP stations: (a) Bishop, CA (040822), mean annual precipitation of 4.9 inches (125 mm); and (b) Quillayute, WA (456858), mean annual precipitation of 101.9 inches (2587 mm). Neither station exhibits appreciable observer bias. Period of record: 1971-2000. Solid curve is the fitted gamma function.

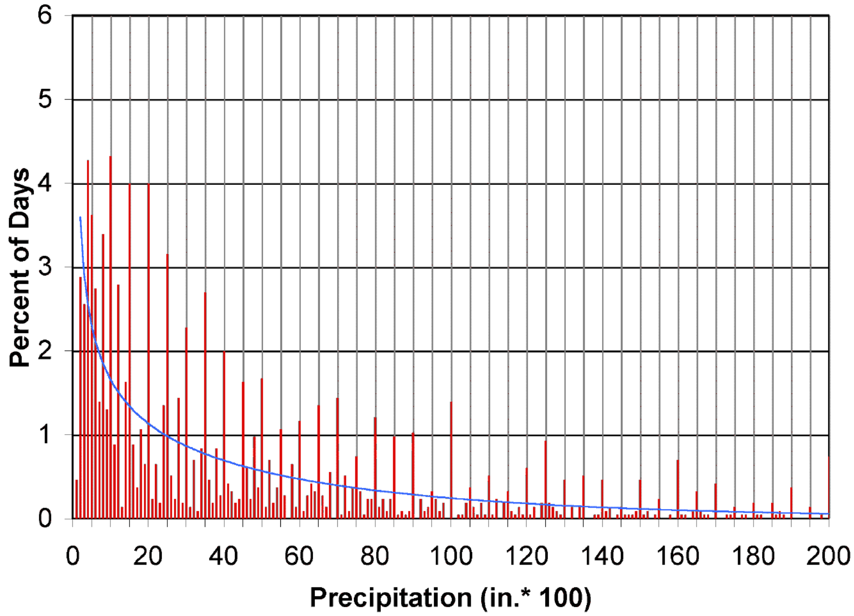


(a)

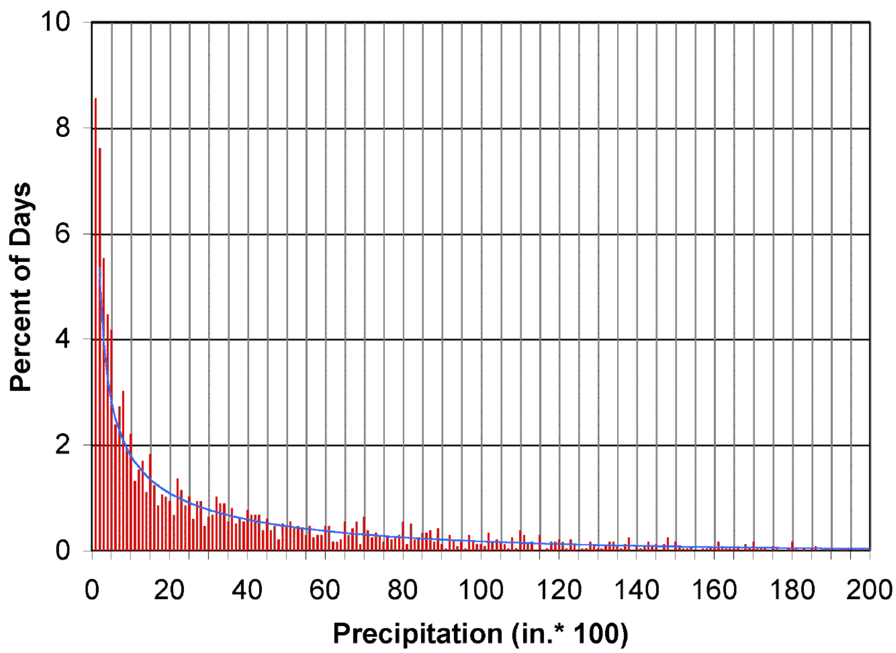


(b)

Figure 3. Percent frequency distribution of daily precipitation of a least 0.01 inch for the period 1971-2000 at COOP stations: (a) Philadelphia 1 WSW, MS (226894), which exhibits a strong under-reporting bias, and some 5/10 bias; and (b) Laurel, MS (224939), approximately 100 km to the south, which exhibits only a slight 5/10 bias. Solid curve is the fitted gamma function.



(a)



(b)

Figure 4. Percent frequency distribution of daily precipitation of a least 0.01 inch for the period 1971-2000 at COOP stations: (a) Purcell 5 SW, OK (347327), which exhibits a under-reporting bias and a strong 5/10 bias; and (b) Watonga, OK (439364), approximately 150 km to the northwest, which exhibits little bias. Period of record: 1971-2000. Solid curve is the fitted gamma function.

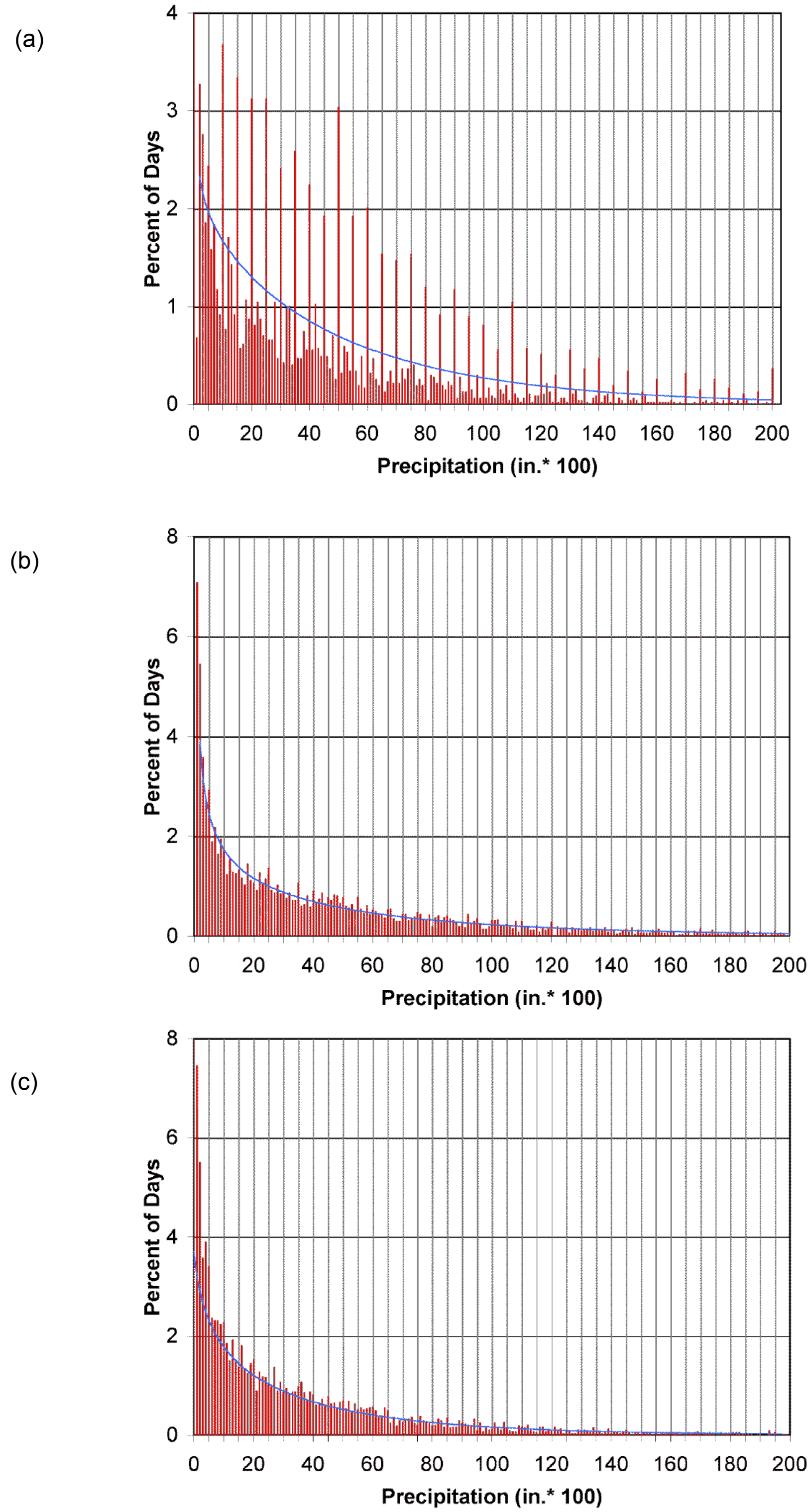


Figure 5. Percent frequency distribution of daily precipitation of a least 0.01 inch for the period 1971-2000 at COOP stations: (a) Cloverdale, OR (351682), which exhibits a under-reporting bias and a very strong 5/10 bias; (b) Otis 2 NE, OR (356366), approximately 20 km to the south, which exhibits no appreciable under-reporting bias, but a small amount of 5/10 bias; and (c) Astoria Airport, OR (350328), approximately 100 km to the north, which exhibits no appreciable bias. Solid curve is the fitted gamma function.

(a)

(b)

(c)

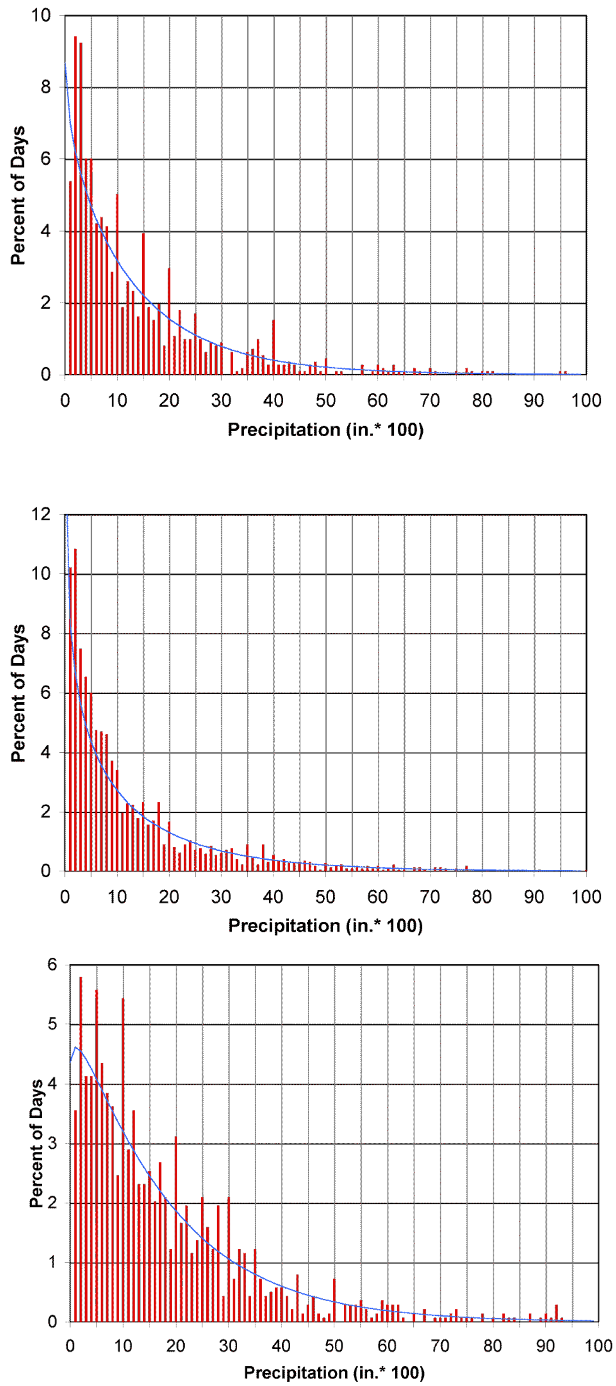


Figure 6. Percent frequency distribution of daily precipitation of a least 0.01 inch at COOP/USHCN station Vale, OR (358797) for the period: (a) 1930-1950, which had a under-reporting bias and strong 5/10 bias; (b) 1950-1980, which was relatively free of bias; and (c) 1980-2005, showing a return to under-reporting bias and 5/10 bias. Solid curves are the fitted gamma functions.

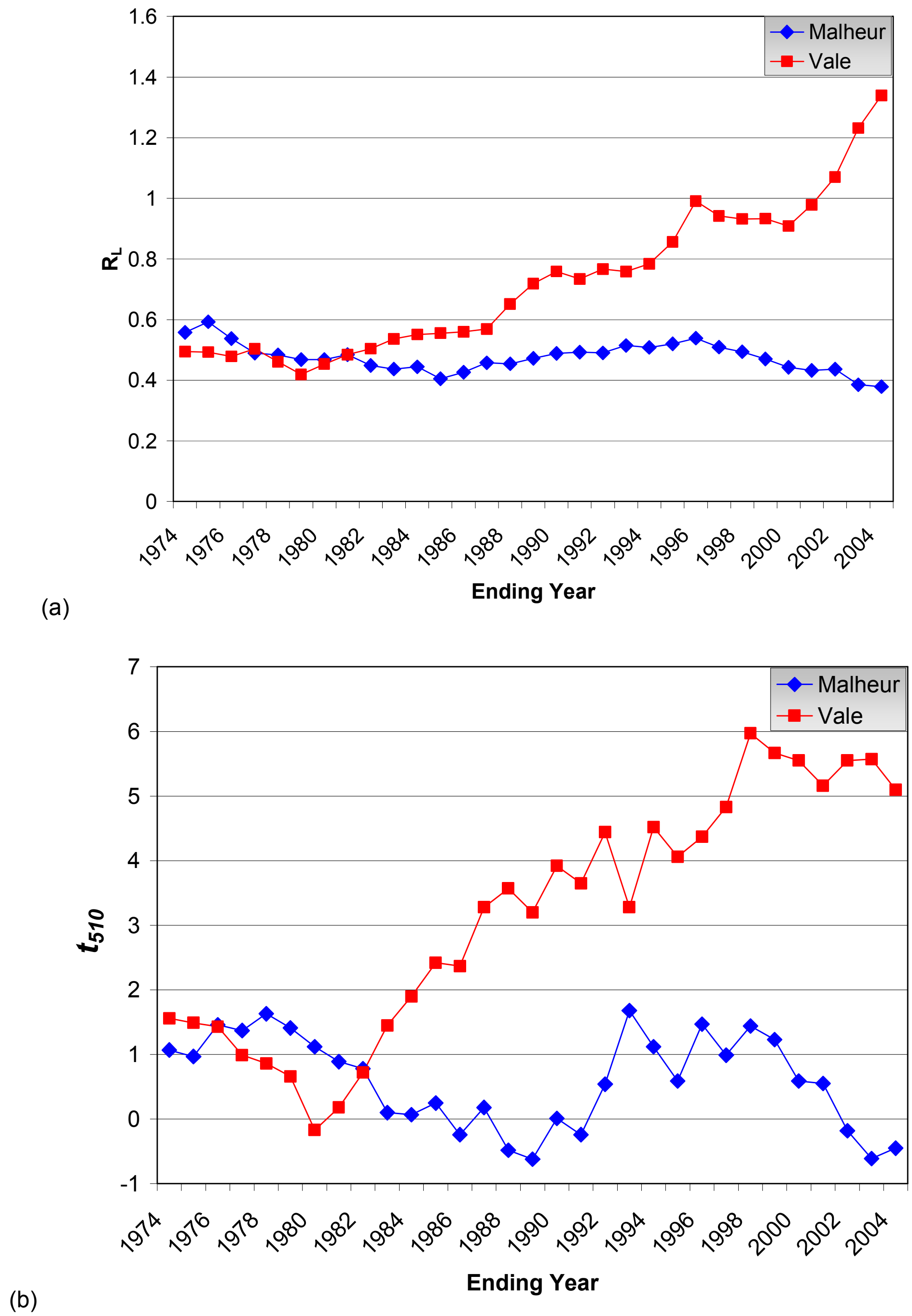


Figure 7. Time series of the 10-year running mean of: (a) the under-reporting bias ratio R_L; and (b) the 5/10 bias maximum t-statistic t_{510}; for COOP stations Vale, OR (358797) and Malheur Branch Experiment Station, OR (355160) for the period 1965-2004. Running means are plotted as year ending, e.g., 1985 represents the period 1976-1985.

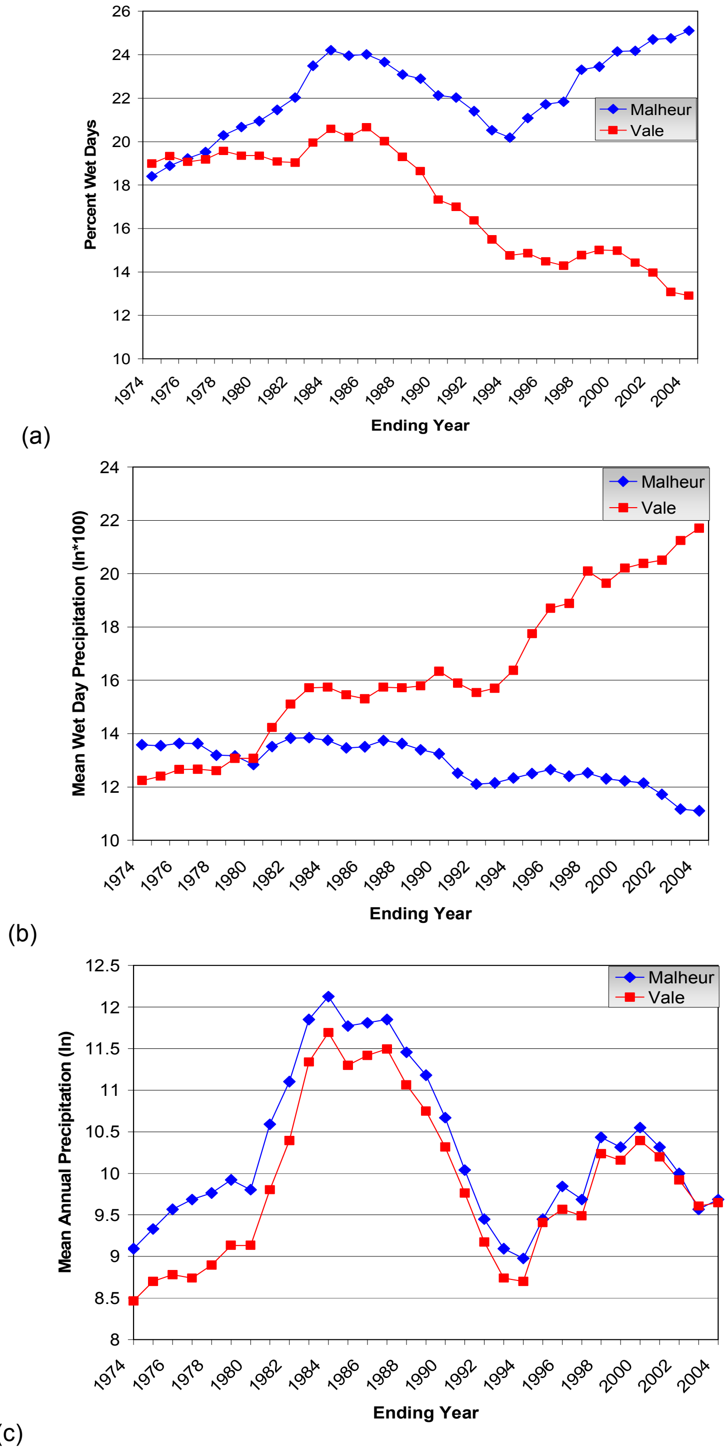
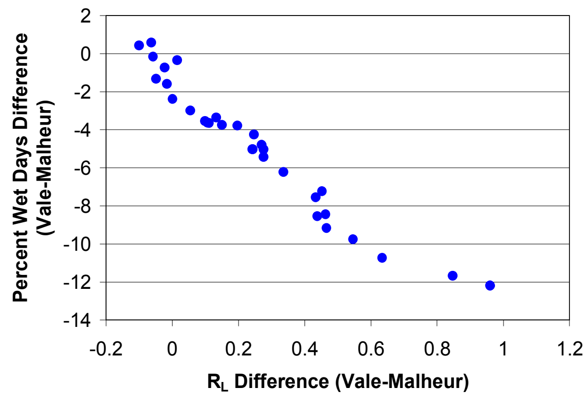
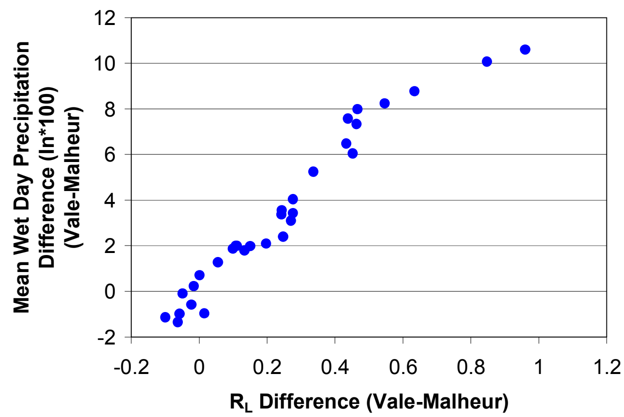


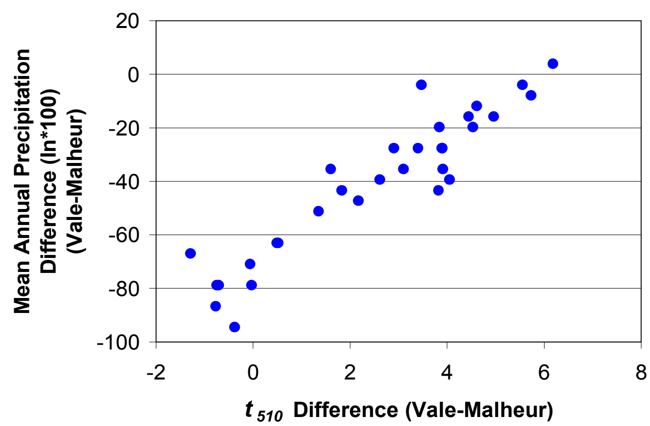
Figure 8. Time series of the 10-year running mean of: (a) the percent of days that are wet (>= 0.01 inch; 0.254 mm); (b) the average precipitation on a wet day; and (c) mean annual precipitation for COOP stations Vale, OR (358797) and Malheur Branch Experiment Station, OR (355160) for the period 1965-2004. Running means are plotted as year ending, e.g., 1985 represents the period 1976-1985.



(a)



(b)



(c)

Figure 9. Scatter plots of differences in the 10-year running means of R_L or t_{510} versus differences in the 10-year running means of commonly-used precipitation statistics for COOP/USHCN stations Vale, OR (358687) and Malheur Branch Experiment Station, OR (355160) for the period 1965-2004: (a) R_L vs. percent of days on which at least 0.01 inch (0.254 mm) or greater was recorded (wet days); (b) R_L vs. average precipitation on a wet day; (c) t_{510} vs. mean annual precipitation.

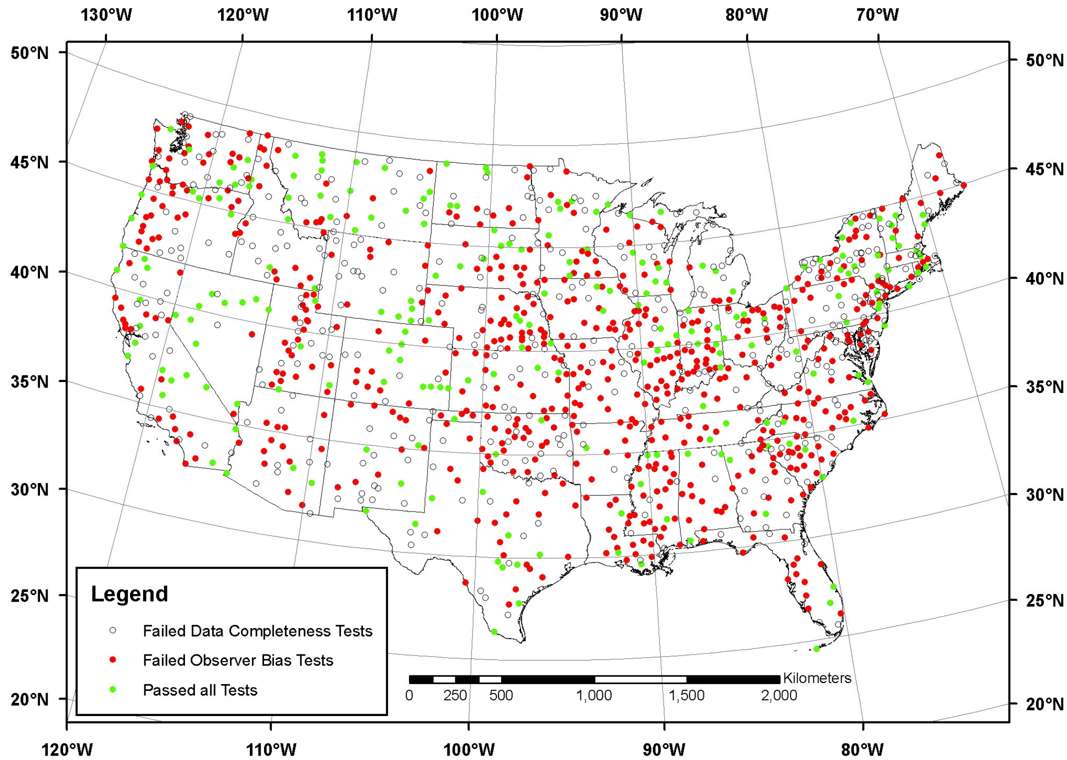


Figure 10. Distribution of USHCN stations passing data completeness and observer bias tests for the period 1971-2000. Only those stations passing the data completeness tests were subjected to the observer bias tests.

Table 1. Results of under-reporting bias, and fives bias and tens bias means tests for COOP stations shown in Figures 4-6, and 9. Period of record is 1971-2000.

Station	Under-reporting bias Ratio (R_L)	Fives Bias Means Test		Tens Bias Means Test	
		T-Stat	P-Value	T-Stat	P-Value
Bishop, CA (040822)	0.31	0.05	0.480	0.25	0.402
Quillayute, WA (456858)	0.44	-1.26	0.106	-1.43	0.079
Philadelphia 1 WSW, MS (226894)	3.57†	4.13	0.000*	4.33	0.000*
Laurel, MS (224939)	0.44	1.97	0.026	0.91	0.184
Purcell, OK (347327)	0.95†	10.79	0.000*	9.37	0.000*
Watonga, OK (349364)	0.40	1.43	0.078	0.58	0.283
Cloverdale, OR (351682)	0.84†	19.18	0.000*	17.03	0.000*
Otis 2 NE, OR (356366)	0.43	3.49	0.000*	0.99	0.163
Astoria Airport, OR (350328)	0.48	1.32	0.096	1.08	0.142
Vale, OR (358797)	0.63†	4.43	0.000*	4.63	0.000*
Malheur Branch Exp. Sta., OR (355160)	0.47	1.21	0.116	-0.06	0.476

* Failed means test at alpha = 0.01
† Failed under-reporting bias test at threshold = 0.6

Table 2. Number and percent of COOP stations that pass data completeness and observer bias tests for the period 1971-2000. USHCN stations (a subset of the "All COOP" stations) are broken out separately. Only 6 percent of all COOP and 18 percent of USHCN stations passed all tests.

	All COOP		USHCN	
	Count	Percent	Count	Percent
Total Candidate Stations	12439	100	1221	100
Passed Data Completeness Tests	2807	23	820	67
Passed Neither Bias Test	584	5	149	12
Passed 5/10 Bias Only	92	1	26	2
Passed Under-reporting bias Only	1347	11	424	35
Total Passed All Completeness and Bias Tests	**784**	**6**	**221**	**18**