Observer Bias in Daily Precipitation Measurements at United States Cooperative Network Stations

BY CHRISTOPHER DALY, WAYNE P. GIBSON, GEORGE H. TAYLOR, MATTHEW K. DOGGETT, AND JOSEPH I. SMITH

The vast majority of United States cooperative observers introduce subjective biases into their measurements of daily precipitation.

The Cooperative Observer Program (COOP) was established in the 1890s 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 U.S. 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 U.S. Historical Climate Network (USHCN). The USHCN 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).

AFFILIATIONS: DALY, GIBSON, TAYLOR, DOGGETT, SMITH—Oregon State University, Corvallis, Oregon
CORRESPONDING AUTHOR: Christopher Daly, PRISM Group, Dept. of Geosciences, Oregon State University, Corvallis, OR 97331
E-mail: daly@coas.oregonstate.edu

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-88-6-899

In final form 30 January 2007

Precipitation data (rain and melted snow) are recorded manually every day by over 12,000 COOP observers across the United States. 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-in.-diameter rain gauge, much like the oil level is measured in an automobile (Fig. 1). The National Weather Service COOP Observing Handbook (NOAA–NWS 1989) describes the procedure for measuring precipitation from 8-in. nonrecording gauges as follows:

> 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 in., a large, unlabeled tick mark every 0.05 in., and small, unlabeled tick mark every intervening 0.01 in. (Fig. 2).

Observations of daily precipitation are needed to parameterize stochastic weather simulation models. These models, often called weather generators, are in wide use for a variety of applications. They are easy to use, and have the ability to synthesize long, serially complete time series of weather data that mimic the true climate of a location, which makes them useful in biological and hydrological modeling and climate change investigations, among others (Richardson 1981; Johnson et al. 1996; Katz 1996). Weather generators require input parameters, derived from station observations, which describe the statistical properties of the climate at a location. Many weather generators use 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. Therefore, it is crucial that the relative frequencies and sequences of wet and dry days are accurately portrayed in the input parameters. In addition, precipitation amounts are often 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 mm).

Fig. 1. Corvallis, Oregon, COOP observer Richard Mattix inserting the measuring stick into his rain gauge.

In a recent study, we used COOP precipitation data to extend the work of Johnson et al. (2000) to spatially interpolate input parameters for the Generation of Climate Elements for Multiple Uses (GEM6) weather generator (USDA–ARS 1994). Spatial interpolation of the input parameters would allow daily weather series to be generated at locations where no stations exist. Our goal was to expand the original mapping region from a portion of the Pacific Northwest to the entire conterminous United States.

Initial mapping of some of the precipitation-related GEM6 parameters using daily COOP data produced spatial patterns that were highly discontinuous in space, even on flat terrain away from coastlines. When we investigated the cause of these spatial discrepancies, we found that precipitation data from most of the COOP stations suffered from observer bias; that is, the tendency for the observer to favor or avoid some precipitation values compared to others. Biases included underreporting of daily precipitation amounts of less than 0.05 in. (1.27 mm), and a strong tendency for observers to favor precipitation amounts divisible by 5 and/or 10 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 dataset 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 United States 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 by millimeter equivalents in parenthesis. All other measures are given in standard metric, or MKS, units.

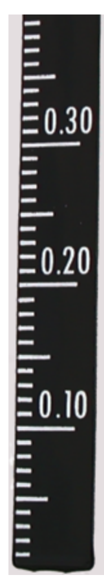


Fig. 2. Standard measuring stick used to record precipitation in a COOP rain gauge.

TYPES OF OBSERVER BIAS AND ASSESSMENT STATISTICS. We discovered two major types of observer bias in our initial investigation: 1) so-called underreporting bias, or underreporting of daily precipitation amounts of less than 0.05 in. (1.27 mm); and 2) so-called 5/10 bias, or overreporting of daily precipitation amounts evenly divisible by 5 and/or 10, such as 0.05, 0.10, 0.15, 0.20, and 0.25 in. (1.27, 2.54, 3.81, 5.08, and 6.35 mm). These two types were usually related; a station with underreporting bias was likely to have a 5/10 bias as well.

We used daily precipitation data from the National Climatic Data Center's (NCDC's) TD3200 dataset (NOAA–NCDC 2006) for this analysis. Each station was subjected to data completeness tests of sufficient rigor to ensure reasonable weather generator 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%) without missing data, and there had to have been at least 26 (85%) complete years within the 1971–2000 period.[1] GEM6 operates on 14-day statistical periods, hence the use of this time block, rather than a monthly time interval.

We devised two kinds of simple statistical tests to detect stations that exhibited one or both observational biases. Our underreporting bias test consisted of calculating the ratio

$$R_L = C_{6-10}/C_{1-5}, \quad (1)$$

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

We assessed possible 5/10 biases 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 compared 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 detrend the frequency histogram. We did this by fitting a gamma distribution to each station's precipitation frequency histogram (Evans et al. 2000). It was not necessary that the gamma function fit the data either precisely, or without bias; rather, the predictions were used only as a way to detrend the frequency distribution.

Because of computational constraints in solving the gamma distribution, predictions become unstable as precipitation approaches zero. Therefore, no frequency predictions were made below 0.03 in. (0.76 mm). (This lower bound has no effect on the detection of underreporting bias, because frequencies were detrended for the 5/10 test only.) In addition, no frequency predictions were made above 1 in. (25.40 mm), because observed frequencies at these precipitation amounts were typically very low.

We calculated the percent difference, or residual (R), between expected and observed frequencies as

$$R = 100 \times (P - O), \quad (2)$$

where P is the predicted frequency (via the gamma function) and O is the observed frequency. We tested the 5s and 10s biases separately. For the fives bias test, the first residual mean ($\overline{R_1}$) was calculated by averaging the residuals over the so-called ones bins, which include all amounts, except those divisible by 5; the second ($\overline{R_5}$) was calculated as the average of all residuals for the so-called five bins, which include only amounts divisible by 5:

$$\overline{R_1} = \frac{\sum_{i=1}^{n_1} R_{1_i}}{n_1}; \quad \overline{R_5} = \frac{\sum_{i=1}^{n_5} R_{5_i}}{n_5}, \quad (3)$$

where n_1 and n_5 are the number of ones and fives bins, respectively, and R_1 and R_5 are residuals, calculated

[1] Tests were conducted using alternative thresholds for data completeness. At 90% completeness, the number of stations available was very low, reducing spatial coverage, and data quality did not improve noticeably compared to the 85% threshold. An 80% threshold admitted more stations, but reductions in data quality became noticeable. The 85% data completeness criterion used here is not dissimilar to those used by the NCDC and the World Meteorological Organization (WMO). When developing monthly precipitation statistics in its TD3220 dataset, the NCDC calculated, but flagged, monthly precipitation totals with one to nine missing days (70%–97% completion), and did not calculate total precipitation for months with more than nine missing days (NOAA–NCDC 2003). WMO guidelines for computing 30-yr normals defined a missing month as having 5 or more consecutive daily values missing (83% completion), or a total of 11 or more missing daily values in the month (63% completion; WMO 1989).

from Eq. (2) that fall into the ones and fives bins, respectively. The tens mean was calculated similarly, but using only bins divisible by 10.

We then used a t-test for comparing the means for small samples, testing the hypothesis that $\overline{R}_1$ and $\overline{R}_5$ were not equal.[2] A t statistic was calculated:

$$t = (\overline{R}_1 - \overline{R}_5)/[s^2\ (n_1{-}1 + n_5{-}2)]^{0.5}, \tag{4}$$

where s^2 is the pooled variance. A two-tailed rejection region $t < -t_{\acute{\alpha}/2}$, $t > t_{\acute{\alpha}/2}$ was established, where $\acute{\alpha}$ is the alpha, or significance, level for the test. Although our main interest was in cases for which the fives-bin residuals were significantly greater than the ones-bin residuals, we applied a two-tailed t-test in case there were situations for which the opposite was true, suggesting an avoidance of the fives bins. This occurred only rarely. We followed a similar procedure for the tens test.

The underreporting 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 underreporting bias test were set, and stations that failed any of the tests were removed from the dataset. 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 (R_L) for the underreporting bias test was 0.60. These threshold values are assumed throughout this paper.

[2] The t-test assumes that the distributions of the samples being compared are generally normally distributed. We applied the Lilliefors test for normality to test this assumption, and found that it was occasionally violated in cases for which the gamma distribution did not fit well or the station was severely biased. As an alternative to the t-test, we applied the Mann–Whitney Wilcoxon rank-sum test, which does not assume normality, and obtained very similar results to those using the t-test. This gave us confidence that the overall sample distributions were sufficiently normally distributed for the t-test.

EXAMPLES OF OBSERVER BIAS. Figure 3 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, California (COOP ID 040822), is a desert site with a mean annual precipitation of 4.90 in. (125 mm), and Quillayute, Washington (456858), is a coastal rainforest site with a mean annual precipitation of 101.90 in. (2588 mm). Despite representing extremely different precipitation regimes, the histograms have a remarkably similar shape. Both stations exhibit a maximum frequency at 0.01 in. (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

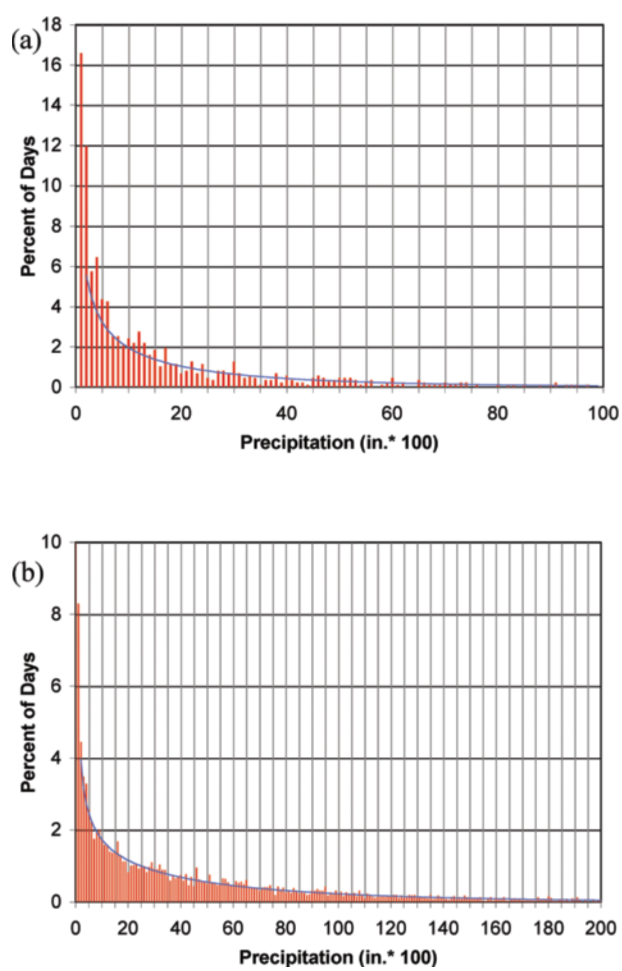
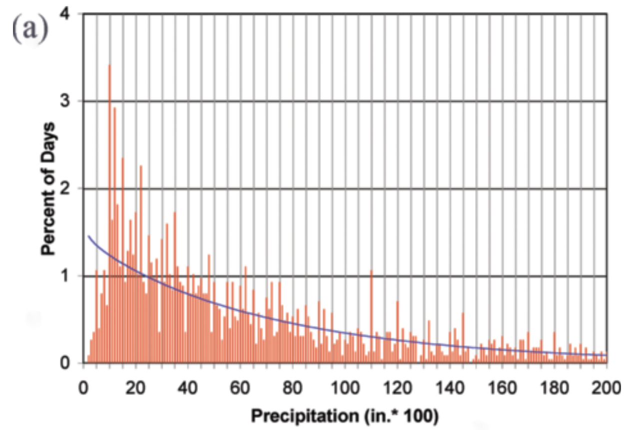


FIG. 3. Percent frequency distribution of daily precipitation of at least 0.01 in. for the period 1971–2000 at COOP stations: (a) Bishop, CA (040822), mean annual precipitation of 4.9 in. (125 mm); and (b) Quillayute, WA (456858), mean annual precipitation of 101.9 in. (2587 mm). Neither station exhibits appreciable observer bias. Solid curve is the fitted gamma function.

as the Quillayute histogram. Generally, the more precipitation events included, the smoother the appearance of the frequency histogram; at least 10 yr 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 4–6 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, Mississippi (COOP ID 226894; Fig. 4a), suffers from a considerable underreporting bias (Table 1). Instead of the expected decreasing frequency trend between 0.01 in. (0.25 mm) and 0.10 in. (2.54 mm), Philadelphia exhibits a sharply increasing frequency trend, with virtually no observations of 0.01 in. (0.25 mm) during the 30-yr period. This station also suffers from 5/10 bias (Table 1), with several frequency spikes at amounts divisible by 10. The number of observations of 0.10 in. (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 in. (7.40, 9.90, and



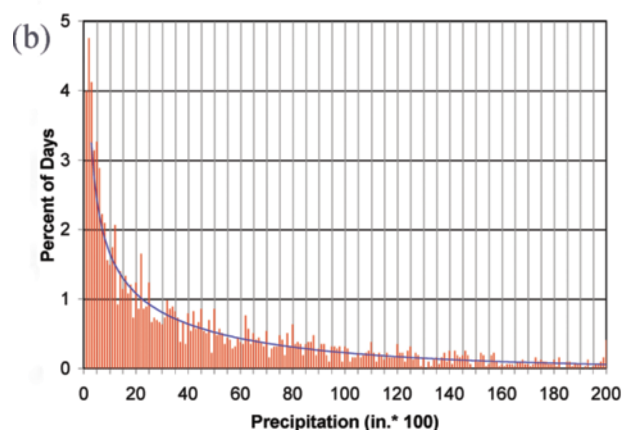


Fig. 4. Percent frequency distribution of daily precipitation of at least 0.01 in. for the period 1971–2000 at COOP stations: (a) Philadelphia 1 WSW, MS (226894), which exhibits a strong underreporting bias and a small 5/10 bias; and (b) Laurel, MS (224939), approximately 100 km to the south, which exhibits only a slight 5/10 bias. Solid curve is the fitted gamma function.

Table 1. Results of underreporting bias, and 5s bias and 10s bias means tests for COOP stations shown in Figs. 4–6, and 9. Period of record is 1971–2000.

Station	Underreporting bias ratio (R_L)	5s bias means test		10s bias means test	
		t statistic	*p* value	*t* statistic	*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[b]	4.13	0.000[a]	4.33	0.000[a]
Laurel, MS (224939)	0.44	1.97	0.026	0.91	0.184
Purcell, OK (347327)	0.95[b]	10.79	0.000[a]	9.37	0.000[a]
Watonga, OK (349364)	0.40	1.43	0.078	0.58	0.283
Cloverdale, OR (351682)	0.84[b]	19.18	0.000[a]	17.03	0.000[a]
Otis 2 NE, OR (356366)	0.43	3.49	0.000[a]	0.99	0.163
Vale, OR (358797)	0.63[b]	4.43	0.000[a]	4.63	0.000[a]
Malheur Branch Exp. Sta., OR (355160)	0.47	1.21	0.116	−0.06	0.476

[a]Failed means test at alpha = 0.01.

[b]Failed underreporting bias test at threshold = 0.60.

12.40 mm, respectively), possibly rounding up to the nearest 0.10 in. (2.54 mm). In contrast, the frequency histogram at Laurel, Mississippi (224939), approximately 100 km to the south, shows only a slightly visible 5/10 bias (Fig. 4b) and passed all observer bias tests (Table 1).

Figure 5 compares precipitation frequency histograms from Purcell 5 SW, Oklahoma (347327), and Watonga, Oklahoma (439364), approximately 150 km to the northwest (see Fig. 7). Purcell 5 SW has a considerable underreporting bias, as well as a well-defined 5/10 bias, and fails all observer bias tests (Fig. 5a; Table 1). In contrast, Watonga shows little underreporting bias and perhaps only a slight 5/10 bias, and passed all observer bias tests (Fig. 5b; Table 1). The frequency of daily precipitation values of 0.01–0.03 in. (0.25–0.762 mm) was 2–3 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 25% more total 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 6 compares Cloverdale, Oregon (351682), on the northern Oregon coast, with Otis 2 NE, Oregon (356366), 20 km to the south. Cloverdale (Fig. 6a) exhibits a striking 5/10 bias, with readings divisible by 5 and 10 occurring 3–6 times more often than other amounts. This station failed all observer

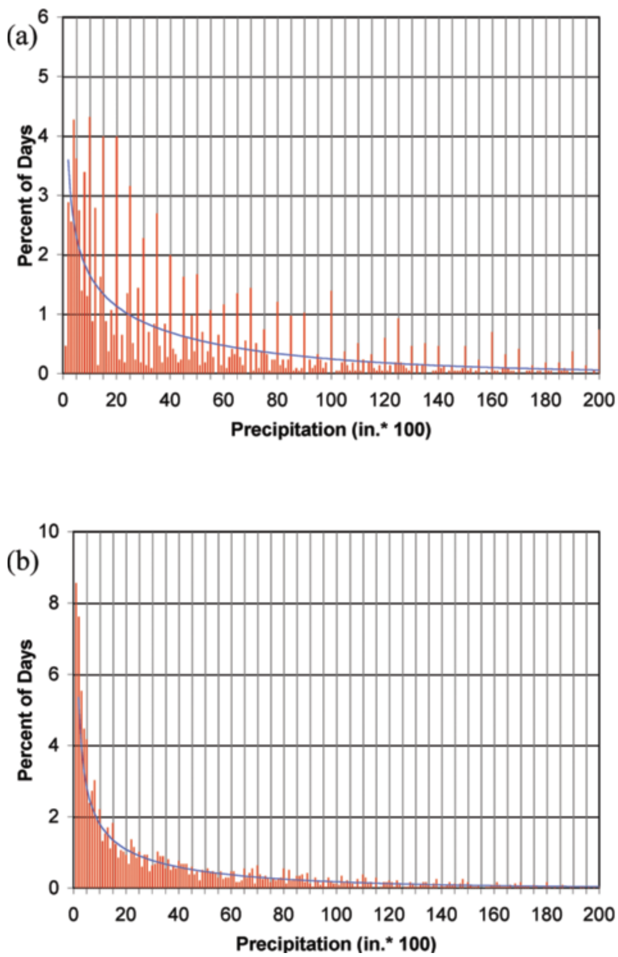


Fig. 5. Percent frequency distribution of daily precipitation of at least 0.01 in. for the period 1971–2000 at COOP stations: (a) Purcell 5 SW, OK (347327), which exhibits an underreporting bias and a strong 5/10 bias; and (b) Watonga, OK (439364), approximately 150 km to the northwest, which exhibits little bias. Solid curve is the fitted gamma function.

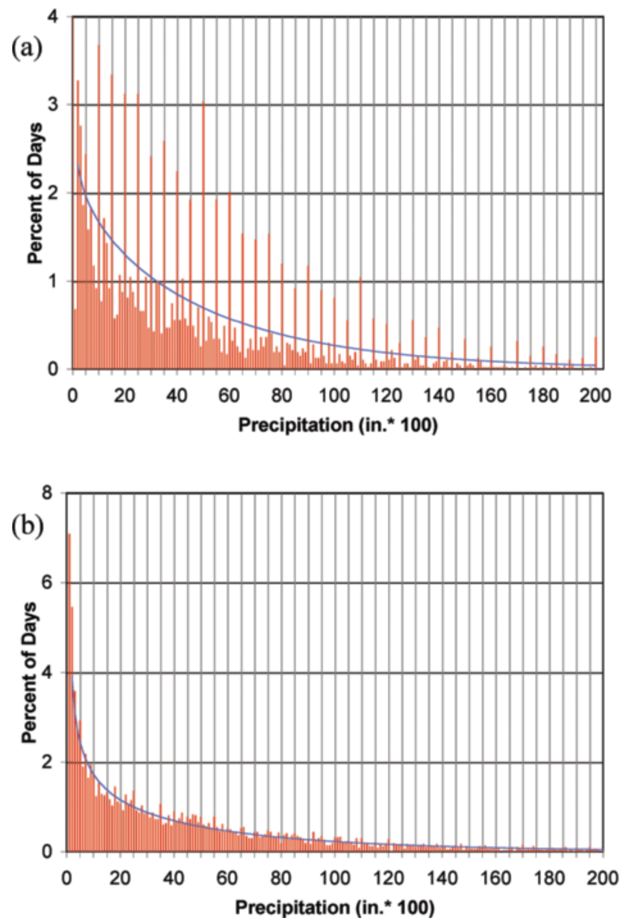


Fig. 6. Percent frequency distribution of daily precipitation of at least 0.01 in. for the period 1971–2000 at COOP stations: (a) Cloverdale, OR (351682), which exhibits an underreporting bias and a very strong 5/10 bias; and (b) Otis NE, OR (356366), approximately 20 km to the south, which exhibits no appreciable underreporting bias, but a small 5/10 bias. Solid curve is the fitted gamma function.

bias tests (Table 1). The frequency histogram for Otis 2 NE (Fig. 6b) 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.

SPATIAL AND TEMPORAL RAMIFICATIONS.

Spatial ramifications. An example of a simple spatial analysis for eastern Oklahoma is shown in Fig. 7. A widely used precipitation statistic, the annual percent of days that were observed to be wet (i.e., days receiving at least 0.01 in. or 0.25 mm) was interpolated with a simple two-dimensional spline fit before and after removal of stations that failed either the underreporting bias or the 5/10 bias tests. Only stations that passed the data completeness tests were considered. The differences in the two maps are striking. The pattern of wet day percentages using all stations is quite complex, with what appears to be a contiguous zone of fewer wet days extending east-to-west through eastern Oklahoma, and more isolated areas of even fewer wet days in the west and south (Fig. 7a). Very few stations in Oklahoma passed all of the observer bias tests, and the map created from these stations shows none of the features just described (Fig. 7b). It shows a general east–west gradient from higher values in the east and lower values in the west, and a tongue of higher values extending westward from Arkansas along the westward extension of the Ouachita Mountains in southeastern Oklahoma.

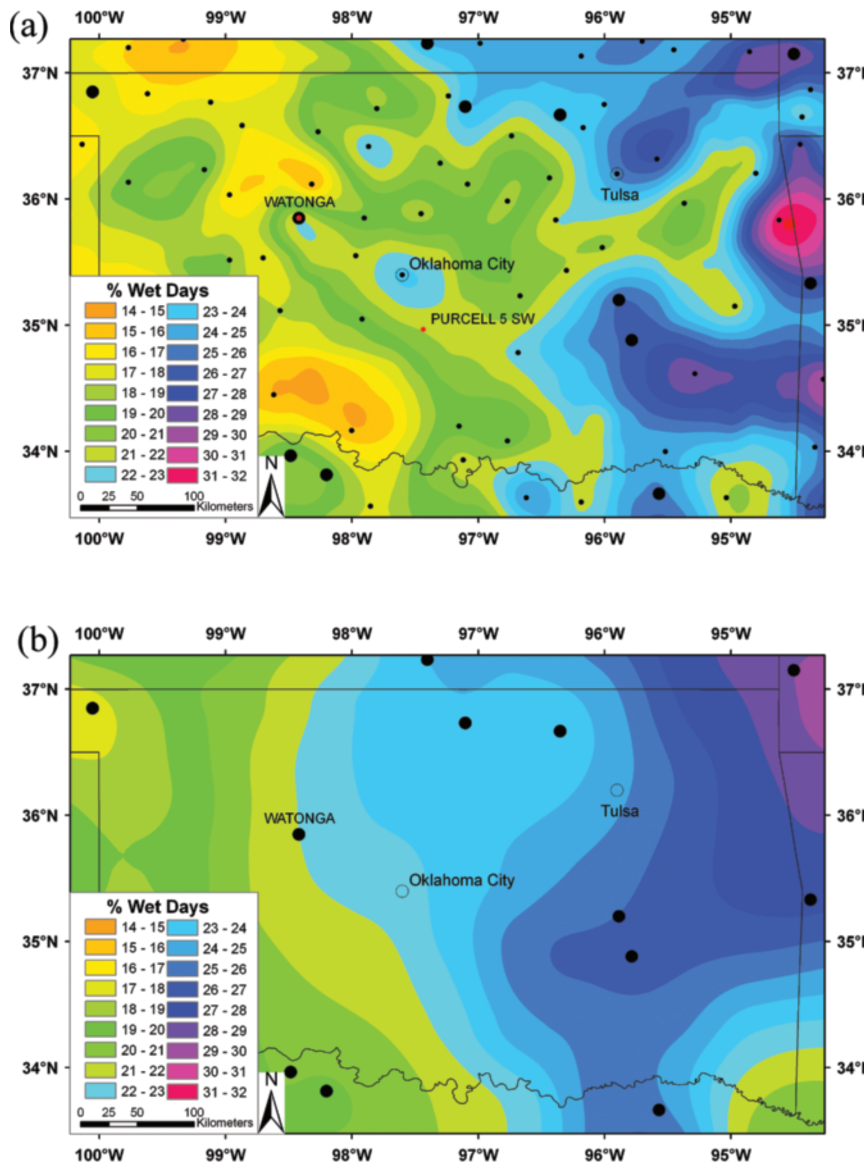


FIG. 7. Spatial distribution of the 1971–2000 mean percent of wet days [days receiving at least 0.01 in. (0.25 mm) of precipitation] over a portion of Oklahoma: (a) using all stations that passed the data completeness tests; and (b) using only stations passing both the data completeness, underreporting bias, and 5/10 bias tests. Stations passing all tests are indicated by large black dots, and those passing only the data completeness tests are shown as small black dots. Purcell 5 SW and Watonga, OK, whose frequency histograms appear in Fig. 5, are marked with red dots.

Perhaps the most interesting aspect of this comparison is that it is not just the spatial patterns that are different, but the actual values, as well. Except for locations that have stations in common in both maps, the percent of wet days was typically about 5% higher on the map created with stations that passed the bias tests than on the all-station map. As was discussed in the comparison of frequency histograms for Purcell 5 SW and Watonga, Oklahoma (Fig. 5), this appears to be due to an unusually high frequency of zero precipitation observations at underreporting-biased stations, which may be a result of higher thresholds for inconsequential precipitation for some observers than for others.

It is frankly difficult to accept the removal of so many stations from a spatial analysis, and we cannot help but wonder if any real climatic features were eliminated by removing so many stations. However, the frequency histograms such as those in Fig. 5 are a reminder of how poor and misleading the distribution of precipitation observations can be at many COOP stations.

Temporal ramifications. The severity of observer bias was often found to vary over time. Consider USHCN station Vale, Oregon (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 (Fig. 8). Vale failed all three observer bias tests, while Malheur

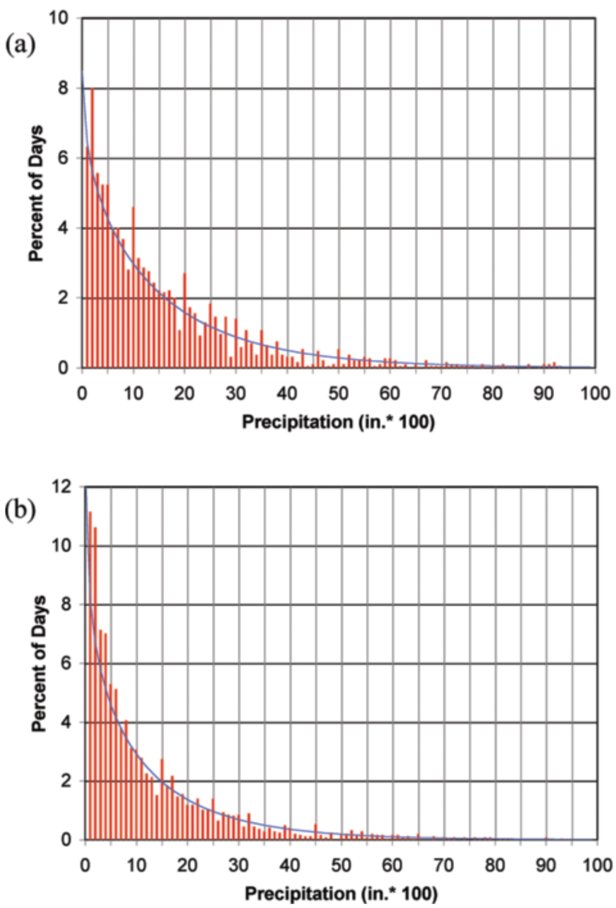


Fig. 8. Percent frequency distribution of daily precipitation of at least 0.01 in. for the period 1971–2000 at COOP stations: (a) Vale, OR (358797), which exhibits a subtle underreporting bias and a more obvious 5/10 bias; and (b) Malheur Branch Experiment Station (355160), approximately 20 km to the east, which exhibits a small 5/10 bias. Solid curve is the fitted gamma function.

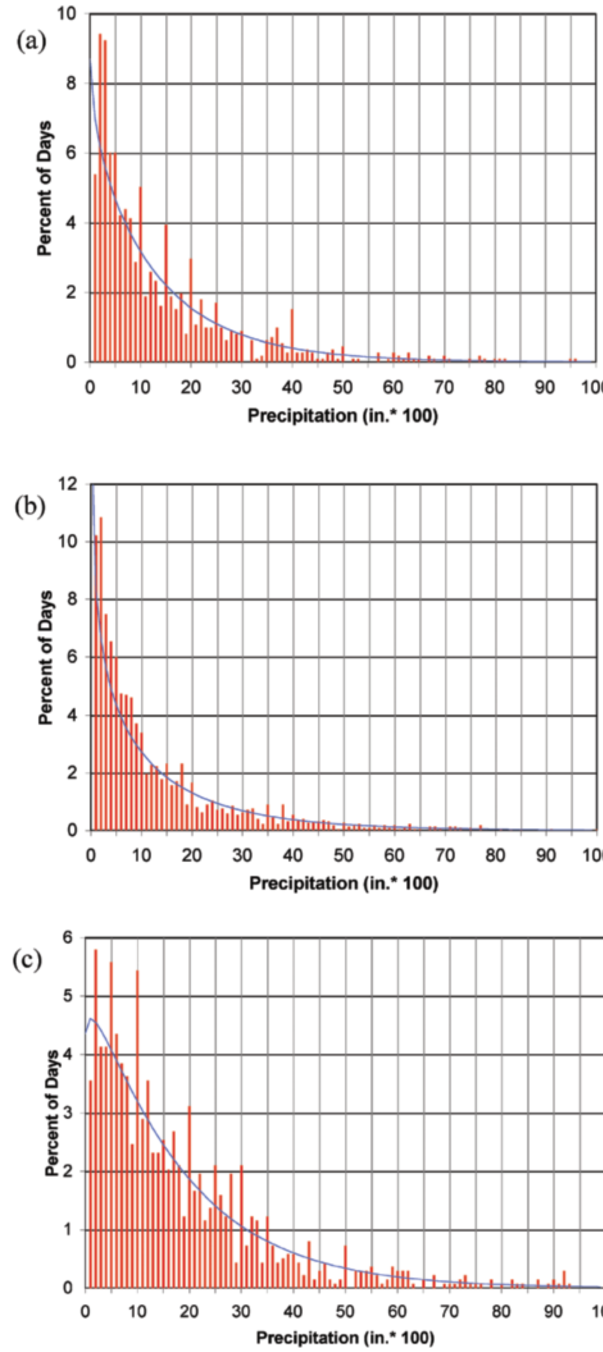


Fig. 9. Percent frequency distribution of daily precipitation of at least 0.01 inch at COOP/USHCN station Vale, OR (358797), for the period (a) 1930–50, which had an underreporting bias and strong 5/10 bias; (b) 1950–80, which was relatively free of bias; and (c) 1980–2005, showing a return to underreporting bias and 5/10 bias. Solid curves are the fitted gamma functions.

passed all three (Table 1). An example of temporal variability in observer bias at Vale is shown in Fig. 9. The period 1930–50 exhibited visible underreporting and 5/10 biases (Fig. 9a), 1950–80 was relatively free of bias (Fig. 9b), and 1980–2005 returned to an underreporting bias and a 5/10 bias (Fig. 9c).

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 10 shows time series of the 10-year running mean of R_L, the underreporting bias ratio, and the maximum of the 5s and 10s bias test t statistics (t_{510}) for Vale and Malheur for the period 1965–2004. The t_{510} statistic reflects the highest t statistic (worst case) of the two 5/10 tests. We used a 10-yr running mean, because at least 10 yr of data are typically required for stable statistical results. We chose the period 1965–2004 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 toward increasing observer bias in both test statistics, while Malheur remained reasonably unbiased throughout the period.

Figure 11 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% during the 10-yr period ending in 1974, but diverged sharply in later years, with Vale trending strongly downward and Malheur slightly upward (Fig. 11a). By the 10-yr period ending in 2004, the percent of wet days at Vale had reached a low of 13%, a 6% drop, while Malheur reported an increase to about 25%, a 7% 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 in. (3.05–5.59 mm) day^{-1}, while Malheur shows a slight drop (Fig. 11b). Trends in mean annual precipitation, a relatively stable precipitation statistic, did not exhibit dramatically different trends, but Vale's value increased relative to Malheur (Fig. 11c).

Relationships between the temporal trends in R_L and t_{510} and those of the three precipitation statistics in Fig. 11 were explored by generating scatterplots of the interstation differences of one versus the other. As seen in Fig. 12, definite relationships exist. As the difference in R_L increased, signaling increased underreporting bias at Vale, there was a strong linear

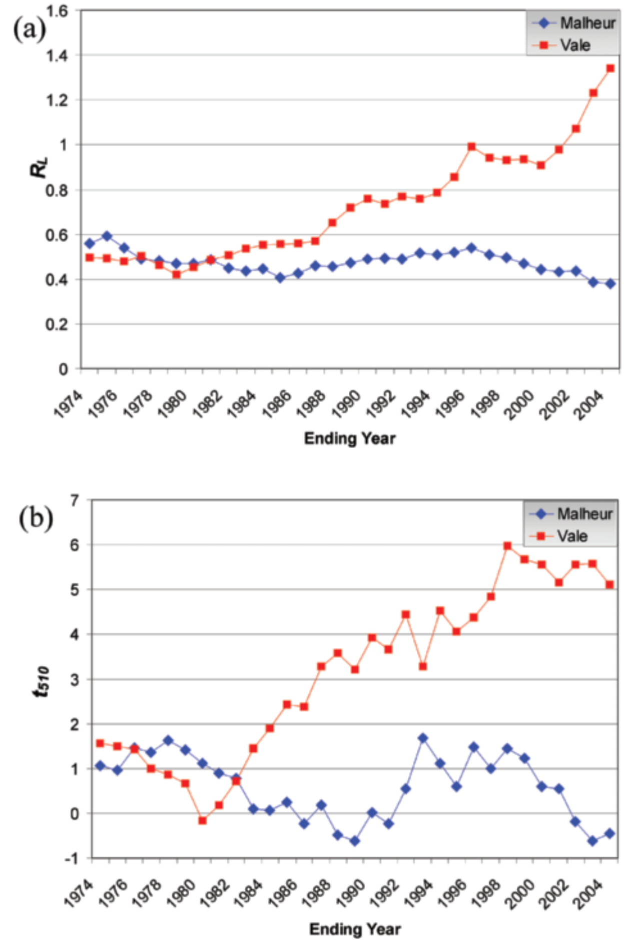


FIG. 10. Time series of the 10-yr running mean of (a) the underreporting bias ratio R_L; and (b) the 5/10 bias maximum t statistic t_{510}; for COOP stations Vale, OR (358797), and Malheur Branch Experiment Station, OR (355160), for the period 1965–2004. Running means are plotted as year ending, e.g., 1985 represents the period 1976–85.

tendency for the number of wet days to decrease compared to Malheur (Fig. 12a). This suggests that the observer increasingly recorded precipitation values of zero or trace on many days, up to 13% more than recorded at Malheur by the 10-yr 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 (Fig. 12b). 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 (Fig. 12c). 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 in. (4.00 mm), up to about 1 in. (25.40 mm). Given that the mean annual precipitation at these two stations was approximately 10 in., the increase in mean annual precipitation at Vale attributable to

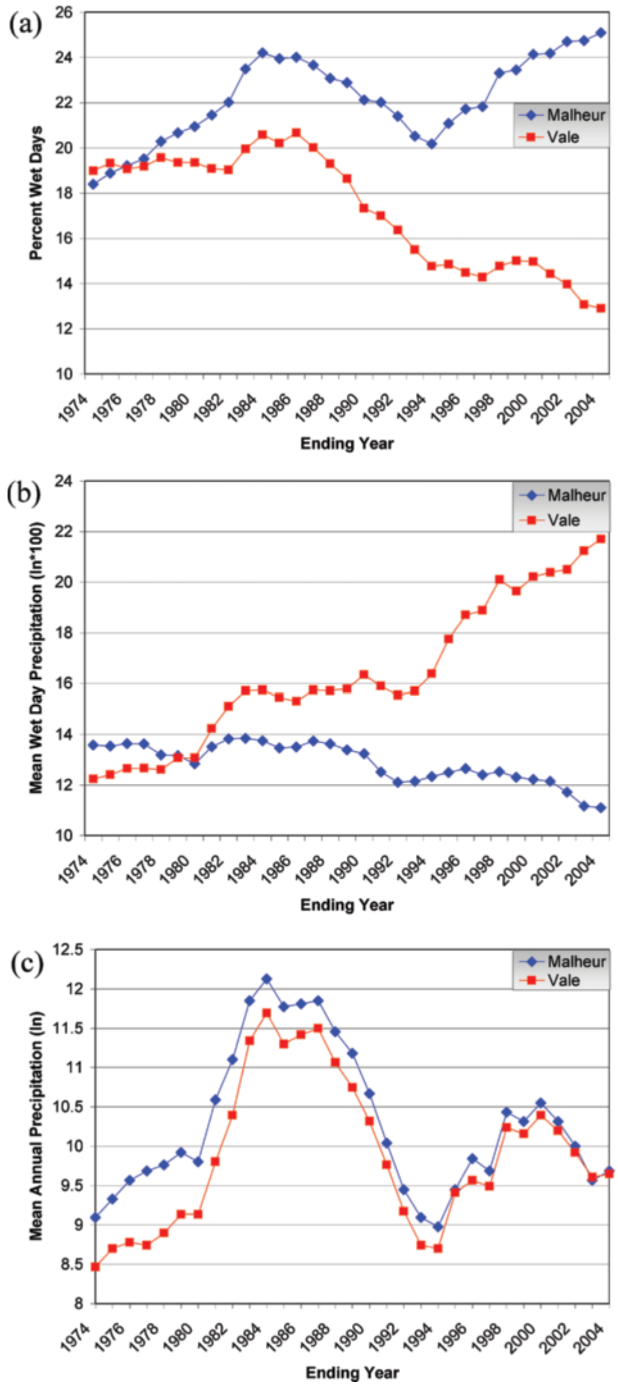


FIG. 11. Time series of the 10-yr running mean of (a) the percent of days that are wet (>=0.01 in.; 0.25 mm); (b) the average precipitation on a wet day; and (c) mean annual precipitation for COOP stations Vale, OR (358797), and Malheur Branch Experiment Station, OR (355160), for the period 1965–2004. Running means are plotted as year ending, e.g., 1985 represents the period 1976–85.

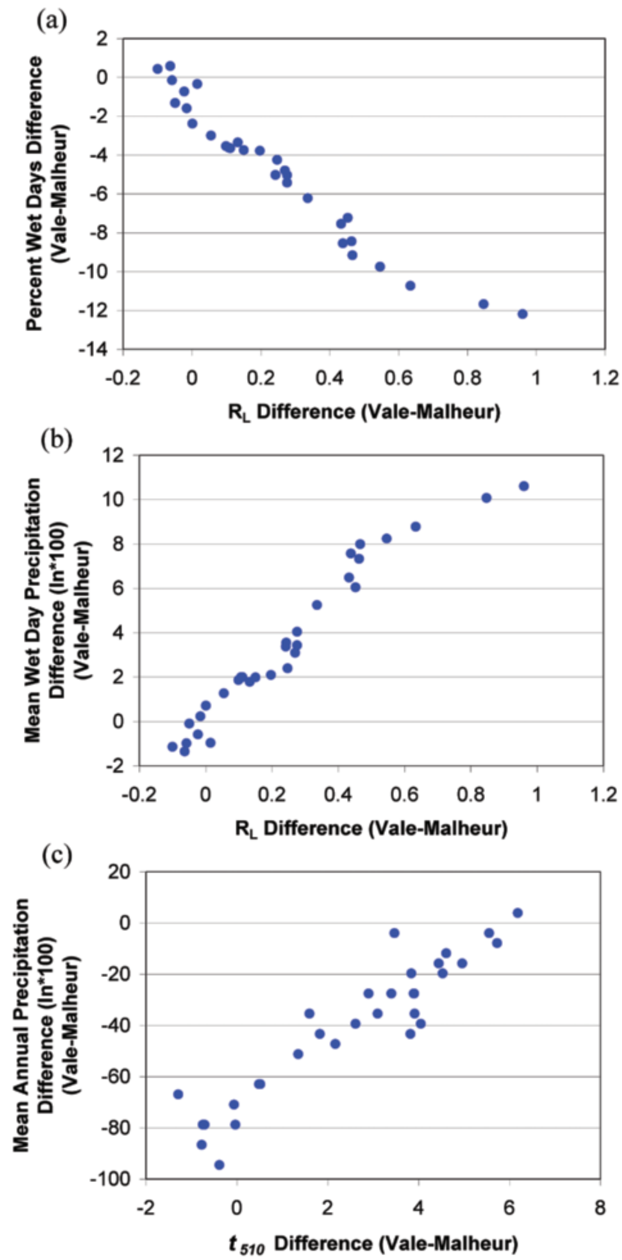


FIG. 12. Scatterplots of differences in the 10-yr running means of R_L or t_{510} versus differences in the 10-yr running means of commonly used precipitation statistics for COOP/USHCN stations Vale, OR (358687), and Malheur Branch Experiment Station, OR (355160), for the period 1965–2004: (a) R_L vs percent of days on which at least 0.01 in. (0.25 mm) or greater was recorded (wet days); (b) R_L vs average precipitation on a wet day; and (c) t_{510} vs mean annual precipitation.

observer bias was about 10%. Reasons for this relationship are not clear, but may have been related to the observer rounding to higher values divisible by 5 and 10.

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% of all COOP and 18% 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 underreporting bias only	1347	11	424	35
Total passed all completeness and bias tests	**784**	**6**	**221**	**18**

SCOPE OF OBSERVER BIAS. To gain perspective on the extent of observer bias in COOP precipitation data across the continental United States, all COOP stations having at least some data within the 1971–2000 period were subjected to the data completeness and observer bias tests. The results, summarized in Table 2 and Fig. 13, were not encouraging. Out of over 12,000 candidate COOP stations, 25% passed the data completeness tests, and of those, 25% passed the observer bias tests, leaving just over 6% of the total, or 784 stations (Table 2). USHCN stations, included in this network partly because of their long and complete records, fared 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%, 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 United States, with all major regions and climate regimes affected (Fig. 13).

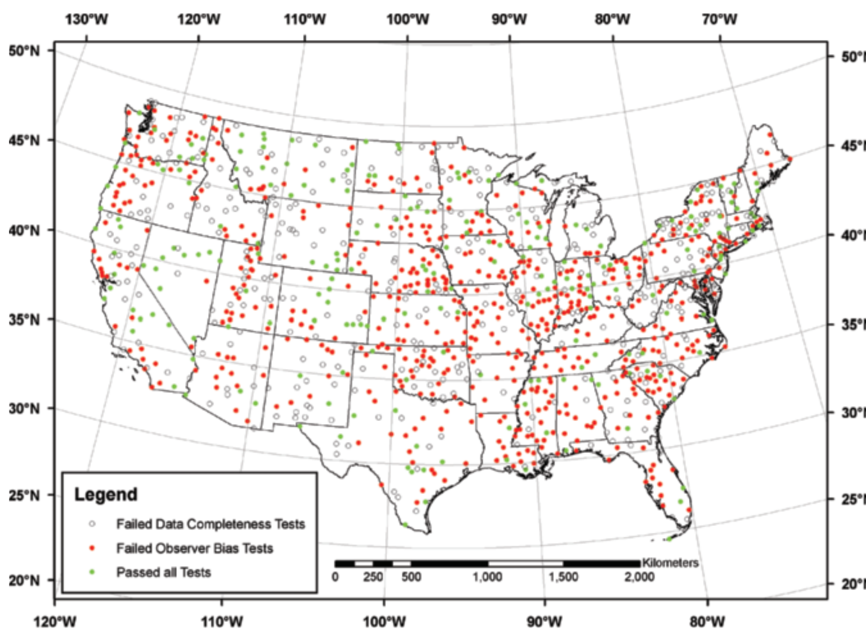


Fig. 13. Distribution of USHCN stations passing data completeness and observer bias tests for the period 1971–2000. Only those stations passing the data completeness tests were subjected to the observer bias tests.

As a check on the reasonableness of the observer bias screening tests, we obtained daily precipitation totals from hourly data at 224 first-order stations from Surface Airways Observation archives for the period 1 July 1996–31 July 2006. To provide sufficient observations for testing, we accepted a station only if it had at least 100 wet days and a total of 3000 nonmissing days. Given that all of these stations employed the Automated Surface Observing System (ASOS) during this period, we would not expect them to suffer from observer bias. All of these stations did pass the observer bias tests, suggesting that these tests, while designed as rough screening devices, were providing reasonable results.

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.[3] Given that the larger and labeled tick marks on the stick represent round values toward which an observer might already gravitate, this additional measuring uncertainty may further motivate observers to choose one of these round numbers.

Another possible cause of a 5/10 bias is the tendency to apportion the precipitation total into two different periods when the observation is not performed at the assigned time (M. Kelsch 2006, personal communication). For example, consider a situation in which it is raining, the observation time is 1700 LST, and the observer is not able to read the gauge until 1800 LST. There is 0.28 in. in the gauge. The observer knows the rain started a little before 1500 LST and has been fairly steady since then. Therefore, the observer estimates that at least two-thirds of the 0.28 in. had fallen by the observation time of 1700 LST, and the remaining since then. When doing the mental math to apportion the precipitation, there may be a tendency to gravitate to round numbers, so that the 0.28-in. total is split into 0.20 and 0.08 in. In this case, the overall precipitation amount between the 2 days would be accurate to 0.01 in., but the 5/10 bias is introduced when the total is split between two time periods.

There appears to be a strong tendency for observers to favor 0.10 in. (2.54 mm) and underreport lower values, with the occasional exception of 0.05 in. (1.27 mm). It is possible that many observers do not see the need to take a precipitation measurement if they perceive that inconsequential precipitation had fallen in the last 24 h, or are unaware that any had fallen and do not check the gauge for confirmation. 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 in. (2.54 mm), which may be rounded to that value if the observer reads the measuring stick with a 5/10 bias.

Underreporting of light precipitation might be reduced if the observer could rapidly determine if any rain fell during the observation period. Standard metal 8-in. rain gauges are opaque, and do not allow the observer to determine at a glance if they contain water. A clear plastic gauge mounted nearby would provide a quick assessment of whether a measurement is needed. Plastic may also be better suited than metal for the actual measurement of light precipitation amounts. In a 10-yr comparison of a 4-in. plastic gauge with a standard 8-in. gauge, Doesken (2005) found that the 4-in. gauge consistently collected more precipitation. Much of the difference appeared to occur during very light events, which he attributed to lower wetting and evaporative losses of the plastic surface compared to the metallic surface.

Regardless of the exact reasons, observer bias suggests a lack of understanding of COOP precipitation measurement procedures, an inability or lack of commitment to fully 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? A related step would be to establish vehicles for frequent and effective communication between COOP observers and the National Weather Service, and among the observers themselves. It stands to reason that observers who are actively engaged would be more likely to make accurate observations than those working in relative isolation.

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. National Oceanic and Atmospheric

[3] It appears that the newer measuring stick was part of a batch issued in the early 2000s that drew complaints from observers, and was subsequently discontinued (S. Nelson 2006, personal communication). However, the Corvallis Hyslop observer was not aware of this.

Administration (NOAA) Environmental Real-time Observation Network (NERON) is a new national program designed to accomplish this. In the first phase of NERON, 100 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 5 min 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. Biases associated with automated gauges could prove to be of similar magnitude and complexity to human observer biases, depending on the gauge type.

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 not noted in the standard COOP metadata, and thus can be difficult to track 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. In 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 10 yr (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. A possible outcome from further work would be a gold standard subset of long-term COOP stations exhibiting consistently low observer bias. Such a subset would be valuable for the calculation of means and trends in precipitation statistics most affected by observer bias.

We 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 our database has information. This application can be accessed by the public online at http://www.prismclimate.org/bias/.

ACKNOWLEDGEMENTS. We thank Mike Halbleib for preparing the maps and Eileen Kaspar for careful editing of the manuscript. We thank Matt Kelsch and two anonymous reviewers for their useful and detailed comments. Thanks also go to Greg Johnson for supporting our effort to map weather generator parameters, and allowing us to further investigate what began only as an interesting data quality issue. This work was funded in part by Agreement 68-7482-3-129Y with the USDA-NRCS National Water and Climate Center.

REFERENCES

Doesken, N., 2005: A ten-year comparison of daily precipitation from the 4" diameter clear plastic rain gauge versus the 8" diameter metal standard rain gauge. Preprints, *13th Symp. on Meteorological Observations and Instrumentation,* Savannah, GA, Amer. Meteor. Soc., CD-ROM, 2.2.

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, 84 pp.

Evans, M., N. A. J. Hastings, and J. B. Peacock, 2000: *Statistical Distributions,* Wiley-Interscience, 221 pp.

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,** 389–395.

Johnson, G. L., C. L. Hanson, S. P. Hardegree, and E. B. Ballard, 1996: Stochastic weather simulation: Overview and analysis of two commonly used models. *J. Appl. Meteor.,* **35,** 1878–1896.

——, C. Daly, C. L. Hanson, Y. Y. Lu, and G. H. Taylor. 2000: Spatial variability and interpolation of stochastic weather simulation model parameters. *J. Appl. Meteor.,* **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 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, 389 pp.

Katz, R. W., 1996: Use of conditional stochastic models to generate climate change scenarios. *Climatic Change,* **32,** 237–255.

NOAA–NCDC, cited 2003: Data documentation for data set 3220, summary of the month cooperative. National Climatic Data Center, National Oceanic and Atmospheric Administration. [Available online at http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td3220.pdf.]

——, cited 2006: Daily surface data, DS3200. National Climatic Data Center, National Oceanic and Atmospheric Administration. [Available online at http://cdo.ncdc.noaa.gov/pls/plclimprod/poemain.accessrouter?datasetabbv=SOD.]

NOAA–NWS, 1989: Cooperative station observations. National Weather Service Observing Handbook 2, Observing Systems Branch, Office of Systems Operations, Silver Spring, MD, 83 pp.

——, cited 2006: Building NOAA's environmental real-time observation network: Draft guidelines for NERON program. Silver Spring, MD. [Available online at www.isos.noaa.gov/documents/Files/NERON.PFRD Draft 0.6.pdf.]

Richardson C. W., 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.,* **17,** 182–190.

Snijders, T. A. B., 1986: Interstation correlations and nonstationarity of Burkina Faso rainfall. *J. Climate Appl. Meteor.,* **25,** 524–531.

USDA–ARS, 1994: Microcomputer program for daily weather simulation in the contiguous United States. ARS-114. United States Department of Agriculture, Agricultural Research Service, 38 pp.

Williams, C. N., R. S. Vose, D. R. Easterling, and M. J. Menne, cited 2006: 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. [Available online at cdiac.ornl.gov/ftp/ndp070/ndp070.txt.]

WMO, 1989: Calculation of monthly and annual 30-year standard normals. Prepared by a meeting of experts, Washington, DC, World Meteorological Organization, WCDP 10, WMO-TD 341, 11 pp.