

4-2015

Precipitation Forecasting with Gamma Distribution Models for Gridded Precipitation Events in Eastern Oklahoma and Northwestern Arkansas

Andrew Lang

Steven A. Amburn
National Weather Service

Michael A. Buonaiuto
Oral Roberts University

Follow this and additional works at: http://digitalshowcase.oru.edu/cose_pub

 Part of the [Statistical Models Commons](#)

Recommended Citation

Andrew Lang, Steven A Amburn and Michael A Buonaiuto. "Precipitation Forecasting with Gamma Distribution Models for Gridded Precipitation Events in Eastern Oklahoma and Northwestern Arkansas" *Weather and Forecasting* Vol. 30 Iss. 2 (2015) p. 349 - 367
ISSN: 0882-8156 Available at: <http://works.bepress.com/andrew-sid-lang/23/>

This Article is brought to you for free and open access by the College of Science and Engineering at Digital Showcase. It has been accepted for inclusion in College of Science and Engineering Faculty Research and Scholarship by an authorized administrator of Digital Showcase. For more information, please contact mroberts@oru.edu.

April, 2015

Precipitation Forecasting with Gamma Distribution Models for Gridded Precipitation Events in Eastern Oklahoma and Northwestern Arkansas

Andrew Lang

Steven A Amburn, *national weather service*

Michael A Buonaiuto, *Oral Roberts University*



This work is licensed under a [Creative Commons CC BY-SA International License](https://creativecommons.org/licenses/by-sa/4.0/).

Precipitation Forecasting with Gamma Distribution Models for Gridded Precipitation Events in Eastern Oklahoma and Northwestern Arkansas*

STEVEN A. AMBURN

National Weather Service, Tulsa, Oklahoma

ANDREW S. I. D. LANG AND MICHAEL A. BUONAIUTO

Oral Roberts University, Tulsa, Oklahoma

(Manuscript received 14 May 2014, in final form 24 November 2014)

ABSTRACT

An elegant and easy to implement probabilistic quantitative precipitation forecasting model that can be used to estimate the probability of exceedance (POE) is presented. The model was built using precipitation data collected across eastern Oklahoma and northwestern Arkansas from late 2005 through early 2013. The dataset includes precipitation analyses at 4578 contiguous, 4 km \times 4 km grid cells for 1800 precipitation events of 12 h. The dataset is unique in that the meteorological conditions for each 12-h event were relatively homogeneous when contrasted with single-point data obtained over months or years where the meteorological conditions for each rain event could have varied widely. Grid cells were counted and stratified by precipitation amount in increments of 0.05 in. (1.27 mm) up to 10 in. (254 mm), yielding histograms for each event. POEs were computed from the observed precipitation distributions and compared to POEs computed from two gamma probability density functions ($\alpha = 1$ and $\alpha = 3$). The errors between the observed POEs and gamma-computed POEs ranged between 2% and 10%, depending on the threshold POE selected for the comparison. This accuracy suggests the gamma models could be used to make reasonably accurate estimates of POE, given the percent areal coverage and the mean precipitation over the area. Finally, it is suggested that the areal distribution for each event is representative of the distribution at any point in the area over a large number of similar events. It then follows that the gamma models can be used to make forecasts for the probability of exceedance at a point, given the probability of rain and the expected mean rainfall at that same point.

1. Introduction

In the summer of 2005, a study began to assess the feasibility of making probabilistic quantitative precipitation forecasts (PQPFs) from gamma probability density functions (PDFs). PQPFs were produced as probabilities of exceedance (POEs), that is, the probability that a select rainfall amount will be exceeded over a given period of time at a point. For example, the 12-h probability of precipitation (PoP) forecast by the

National Weather Service (NWS) is a point probability that precipitation will exceed zero rainfall during a 12-h period. The model described here provides a method for computing the probability that higher rainfall amounts might be exceeded. For example, the model might indicate a 30% chance to exceed 1.00 in. (25.4 mm) during one rainfall event, but only a 10% chance to exceed 1.00 in. during a different rainfall event.

The study was conducted using gridded areal rainfall data. POEs were calculated from the observed areal data for 1800 cases (12-h events) and compared to POEs computed for each event from gamma PDFs to see how closely the PDFs fit the observed distributions. The 1800 events were selected to include only those events where the 12-h areal coverage was 10% or greater. There were over 2200 events with areal coverage of 5% and greater and well over 3000 events where at least 0.01 in. of rain was analyzed on at least one 4 km \times 4 km grid cell (hereafter, simply "grid").

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/WAF-D-14-00054.s1>.

Corresponding author address: Steven A. Amburn, National Weather Service, 10159 E. 11th St., Ste. 300, Tulsa, OK 74128.
E-mail: steve.amburn@noaa.gov

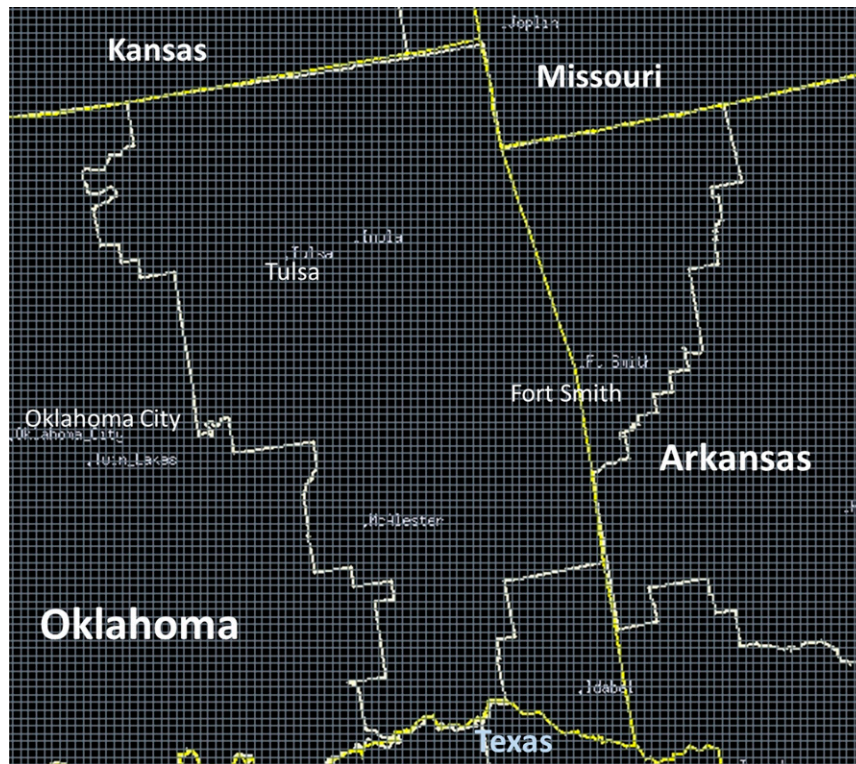


FIG. 1. WFO Tulsa forecast area covers most of eastern Oklahoma and a large portion of northwestern Arkansas. This area is outlined in white. There are 4578 grid cells ($\sim 4 \text{ km} \times \sim 4 \text{ km}$) covering the outlined area.

A number of authors have suggested or shown that precipitation data at a point, collected over time, could be expressed by a gamma distribution (Thom 1958; Wilks 1995). However, there is apparently little if any documentation on whether areal precipitation data from single events at this $4 \text{ km} \times 4 \text{ km}$ scale could also be modeled using gamma distributions. This issue is addressed using areal data from 2005 through early 2013, over the County Warning and Forecast Area (CWFA) of the National Weather Service Weather Forecast Office in Tulsa, Oklahoma (WFO TSA).

Data were obtained from the Arkansas–Red River Basin Forecast Center (ABRFC; ABRFC 2014) routine rainfall analyses over the WFO TSA CWFA and accepted as analyzed with no adjustments. A script was written to produce the unique dataset as text data, which were then stored for WFO TSA to retrieve at its discretion. Although the data are routinely produced, the particular format of the data for this study was not a routine product of the ABRFC and is not a routine from any other river forecast centers. Production of this dataset was gracious compliance to a special request from WFO TSA and is not a routine public product from the ABRFC. Other researchers should be able to construct datasets similar to those used in this study from the

“shape files” of data available on the Advanced Hydrologic Prediction Service website (<http://water.weather.gov/precip/download.php>). For convenience, we have included the data we used in this study as a supplemental data file. The supplemental data file also contains five tabs of examples that illustrate how we used the data to produce our results.

A previous study developed to provide real-time verification for the probability of precipitation (Spaeth 1999) revealed that the ABRFC could provide high-resolution

TABLE 1. Data for 1800 events of 12-h duration from late 2005 through early 2013. Units of the means is inches.

Areal coverage (%)	Conditional mean	Unconditional mean	No. of 12-h events
10.0–19.9	0.094	0.014	354
20.0–29.9	0.132	0.032	262
30.0–39.9	0.136	0.047	224
40.0–49.9	0.156	0.070	158
50.0–59.9	0.208	0.115	148
60.0–69.9	0.269	0.174	156
70.0–79.9	0.256	0.191	142
80.0–89.9	0.310	0.263	119
90.9–99.9	0.432	0.415	163
100	0.577	0.577	74

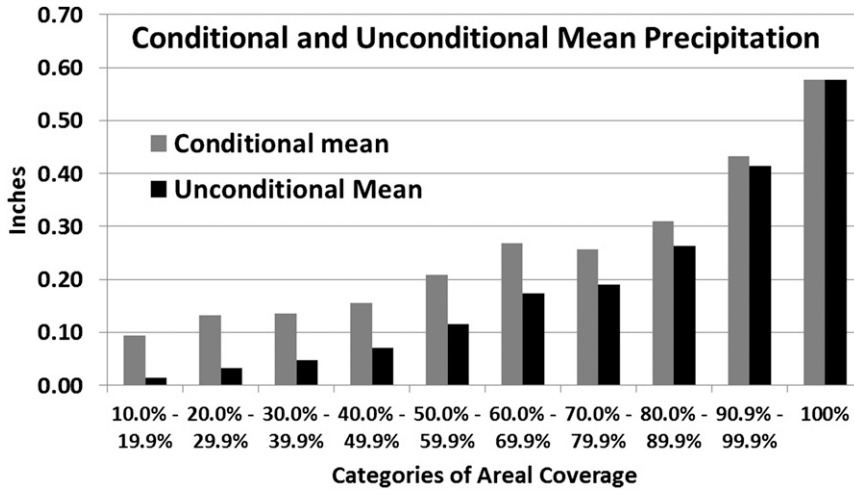


FIG. 2. Graphical depiction of data in Table 1; 1800 events categorized by areal coverage. Both conditional and unconditional mean precipitation amounts are shown by category in inches.

(4 km × 4 km), event-based precipitation data on the areal coverage of precipitation in the TSA forecast area. In addition to coverage, the ABRFC analyses also indicated how much it rained at that resolution. They supplied that data to WFO TSA for this study. To the authors’ knowledge, this is a singular and unique dataset and no other study has used contiguous precipitation data in this manner.

The CWFA mean rainfall for each event was used to create gamma PDFs for each of the 1800 events. POEs from the observed data and those from the gamma computations were compared at specific thresholds. The accuracy was sufficient and consistent enough to suggest that POE forecasts could be made from the gamma

PDFs, based on mean quantitative precipitation forecasts (QPFs). Mean absolute errors (MAEs) were computed for a wide variety of threshold exceedance probabilities to estimate the accuracy of the method. All MAEs were less than 10%. The MAE at the 10% exceedance probability threshold varied from 6% to less than 3%.

The method of computing the probability of exceedance presented here has not been validated over complex terrain but should still work as long as the probability of precipitation and mean QPF are forecast reasonably well. This is a reasonable assumption, as forecasters generally know the effects that local terrain has on their precipitation probabilities and rainfall amount models.

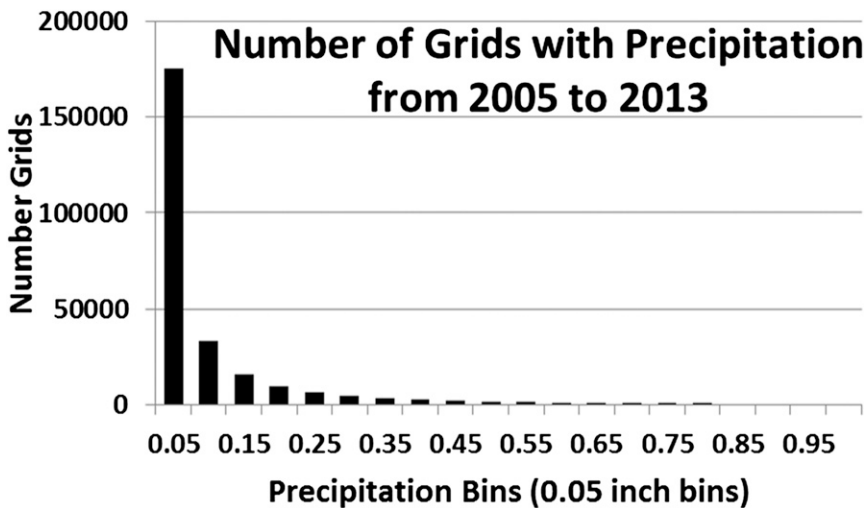


FIG. 3. This histogram shows the sum of grids for the associated precipitation bin increment from fall 2005 through spring 2013 for the WFO TSA forecast area. The chart is truncated at 1 in. (25.4 mm) because of the large number of light precipitation amounts and the scale.

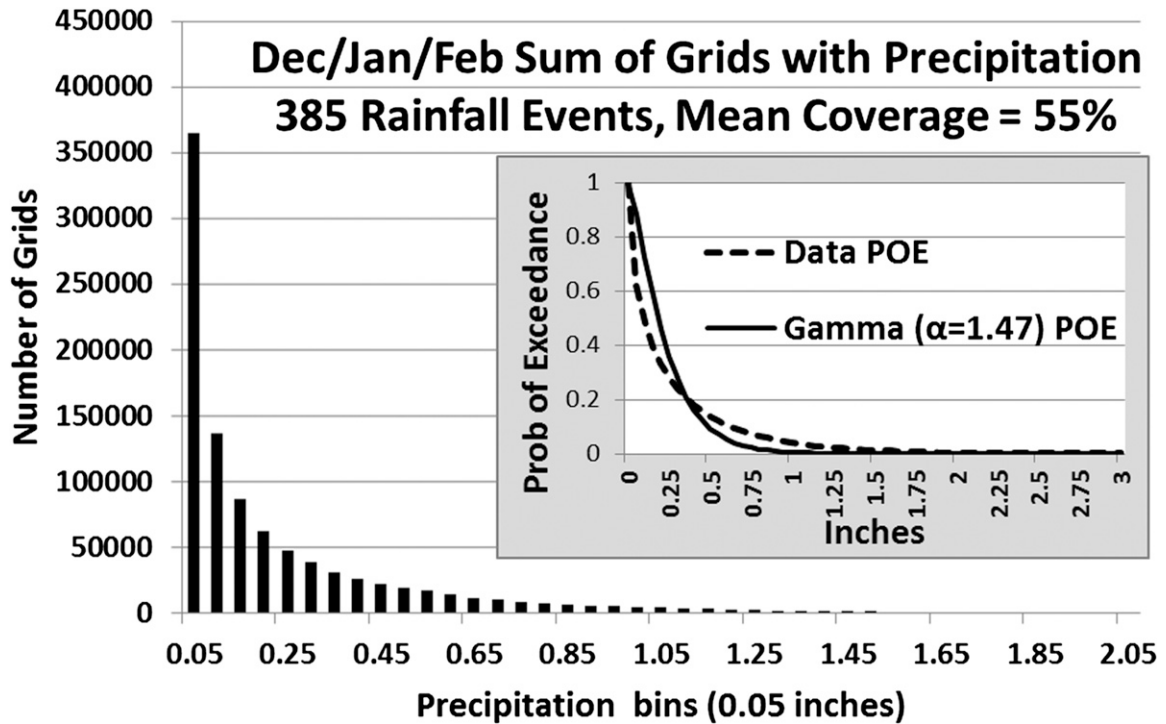


FIG. 4. Histogram of grids with precipitation for meteorological winter for the WFO TSA forecast area. The inset shows the POE computed from the data (solid line) and the gamma-approximated POE using an averaged alpha.

Where precipitation is enhanced by terrain, the mean rainfall will also increase, increasing the POEs. Where terrain enhances the probability that precipitation will occur, the POEs will again increase.

2. Data

The analysis of rainfall data is part of the daily routine at the ABRFC for use in their forecast operations.

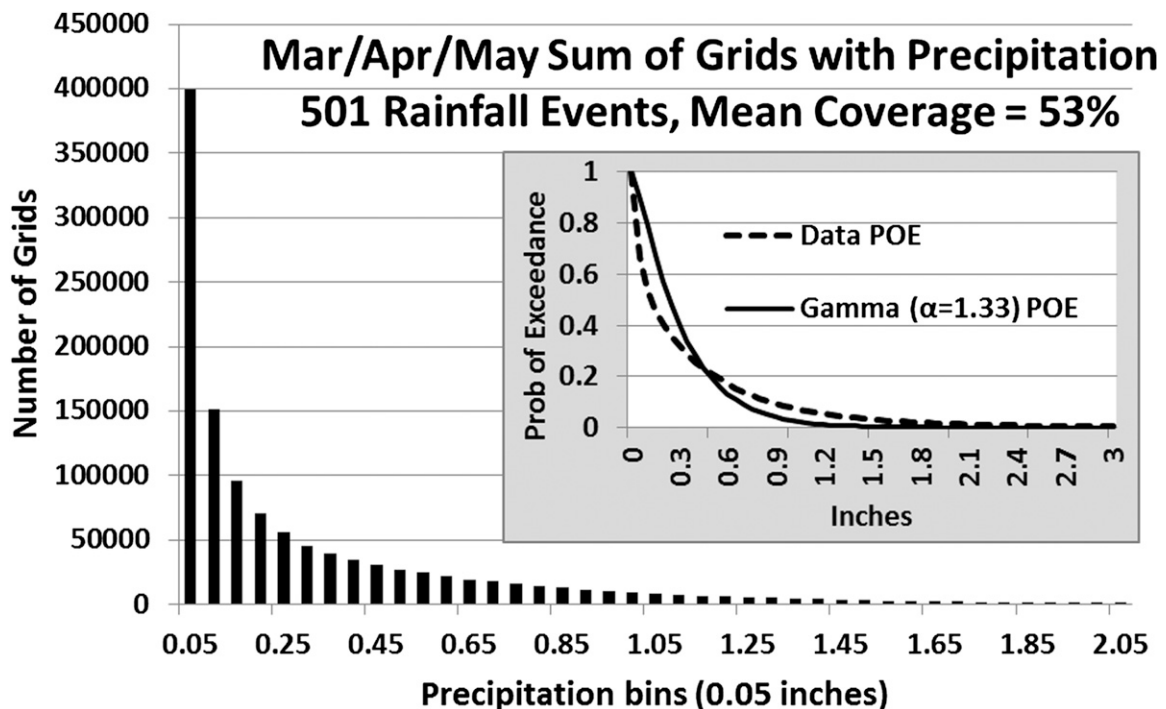


FIG. 5. As in Fig. 4, but for meteorological spring.

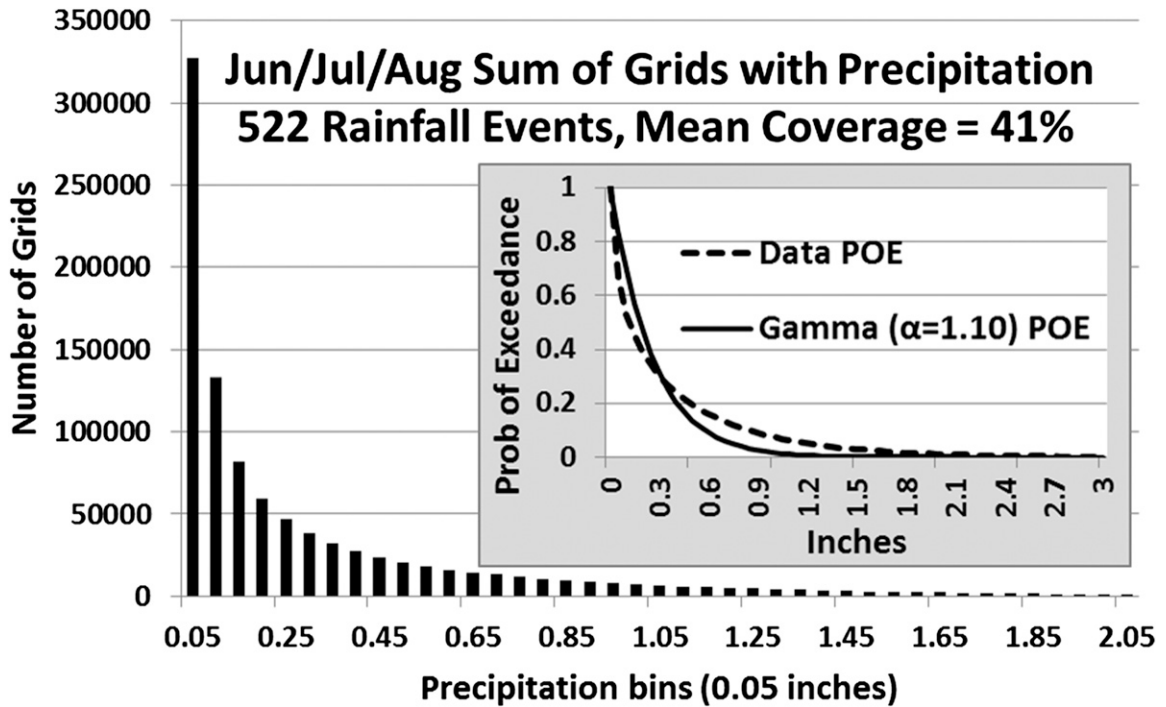


FIG. 6. As in Fig. 4, but for meteorological summer.

Each 4 km × 4 km grid is analyzed for precipitation each hour and then combined for the 12-h analysis period ending at either 0000 or 1200 UTC to match data collection intervals for ABRFC operations. These data

are a subjective analysis that uses a combination of NWS Weather Surveillance Radar-1988 Doppler (WSR-88D) data, available rain gauge data, and human input.

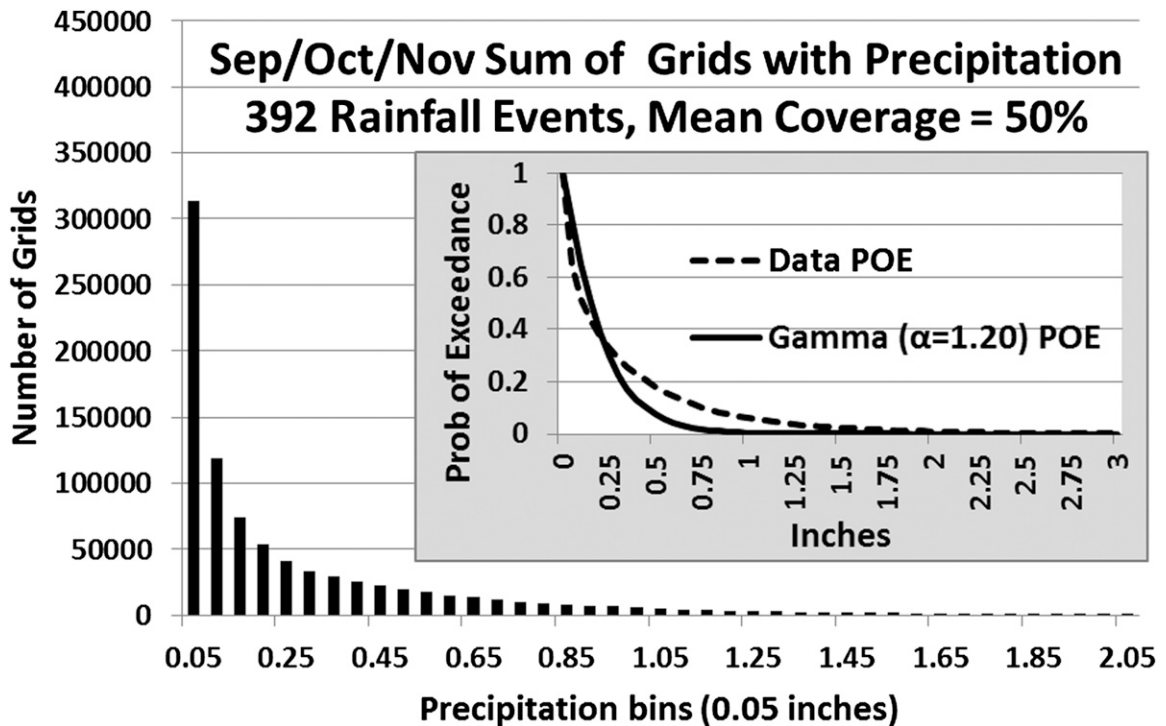


FIG. 7. As in Fig. 4, but for meteorological autumn.

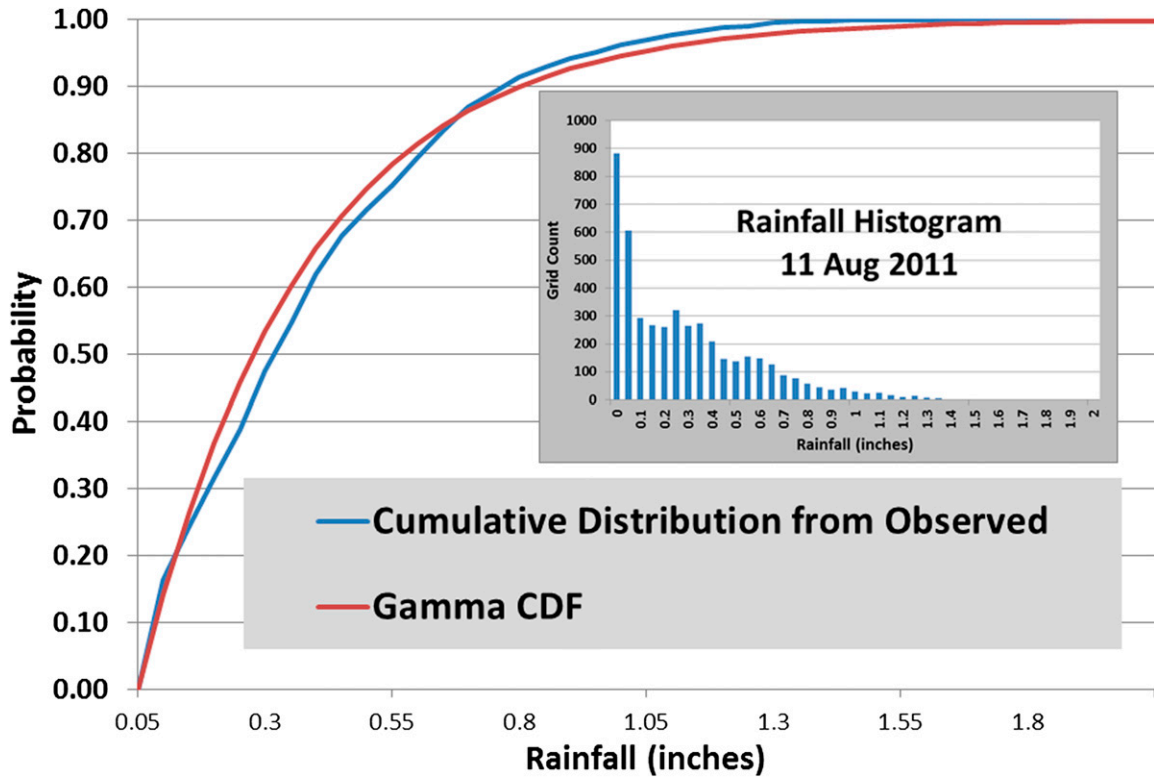


FIG. 8. Cumulative distributions of observed data and the gamma approximation. The inset shows the histogram for the 12-h event. Vertical axis is cumulative probability; horizontal axis is precipitation in inches.

Observed gauge data are used to bias correct WSR-88D data, and the WSR-88D data are used to fill in gaps where observed gauge data are not available. The final result is a contiguous areal analysis of precipitation coverage and amount for each 4 km × 4 km grid (ABRFC 2014). The radar data are corrected for hail contamination, ground clutter, and spurious echoes. Gauge data are corrected for false tips, which may occur from heavy dew or snow in the gauge that melts well after the precipitation event ended. (The data are collected in units of inches and are presented here as such. Occasional reference to millimeters is given when appropriate.)

The grids that cover the TSA CWFA are shown in Fig. 1, encompassing most of eastern Oklahoma and a portion of northwestern Arkansas. Data for each grid are then stratified by precipitation amounts and binned at rainfall increments of 0.05 in. There are 4578 grids (proxy rain gauges) within the TSA CWFA, covering just over 73 000 km² (approximately 28 000 mi²).

The data used in this study were filtered from the original dataset to eliminate those events where less than 10% of the grids were estimated to have received measurable rainfall. This left 1800 events of 12-h duration. Table 1 shows the breakdown of events by percent areal coverage intervals, their associated conditional

and unconditional precipitation means based on the ABRFC quantitative precipitation estimates (QPEs), and the number of events in each interval based on areal coverage. Figure 2 is a graphic depiction of Table 1.

Figure 3 is a histogram of all the data collected from fall 2005 through spring 2013. Figure 3 was truncated at 1.0 in. (25.4 mm) because of scaling considerations resulting from the high count of grids receiving very light precipitation (first bin) and the wide variety of low counts for heavier precipitation. The first bar shows that approximately 175 000 grids received precipitation in the bin representing 0.01–0.05 in. Approximately 33 000 grids received precipitation in the bin from 0.06 through 0.10 in. Data were collected through 10 in. (254 mm), although no events produced 10 in. of precipitation at any grid during a 12-h period.

TABLE 2. Comparison of other distributions considered in modeling the data.

PDF	AIC
Poisson	2307
Negative binomial	260
Zero inflated (negative binomial)	260
Hurdle (negative binomial)	246
Cauchy	222

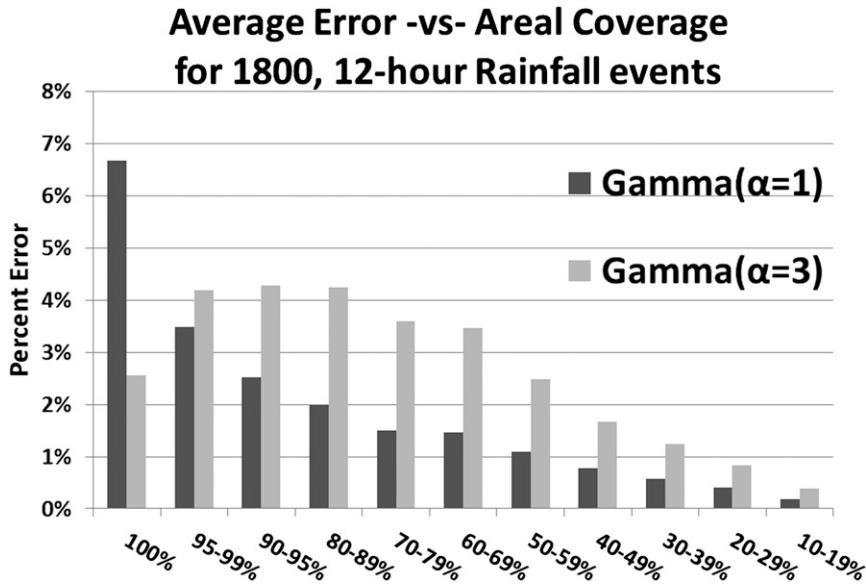


FIG. 9. Comparison of percent error vs areal coverage. For events with 100% areal coverage, the gamma PDF where $\alpha = 1$ had an average error of over 6.5%. However, the gamma PDF where $\alpha = 3$ had an average error of only about 2.5%. The “crossover” where one outperformed the other occurred between 90% and 95% areal coverage.

Figures 4–7 are combined histograms for meteorological seasons with precipitation bins through 2.05 in. Again, the data shown only include events where there was 10% or more areal coverage of reportable precipitation (0.01 in. or more). A climatology of seasonal exceedance probabilities could be easily computed from

the data in Figs. 4–7. A similar process was completed for 108 single stations in the conterminous United States (Jorgensen et al. 1969) using 15 years of data. The seasonal POEs they calculated, which required considerable manual effort, are very similar to seasonal POEs computed from this current study.

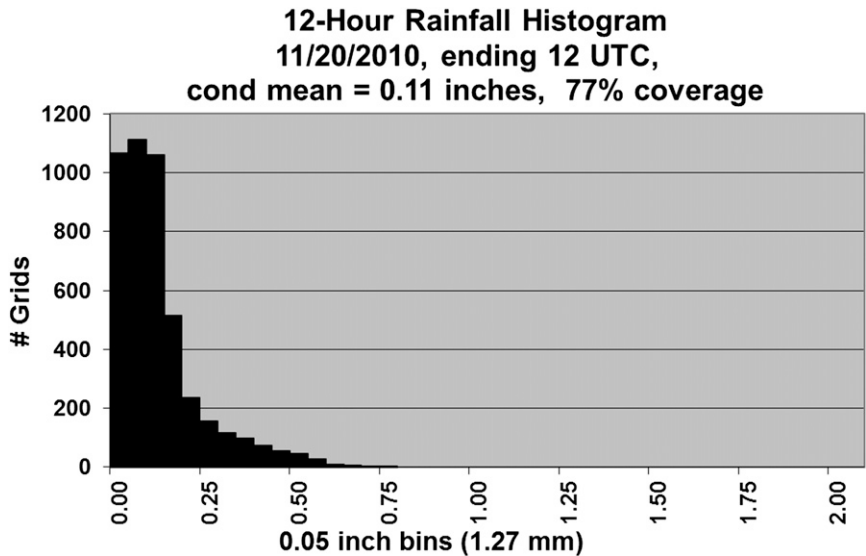


FIG. 10. This precipitation histogram for 20 Nov 2010 is for the 12-h period ending at 1200 UTC. Vertical axis shows the number of grids within the labeled precipitation grouping, starting with nearly 1300 grids from 0.01 through 0.05 in. The histogram is truncated where no additional precipitation occurred. Similar histograms can be produced for each of the 1800 events with 12-h period where precipitation occurred.

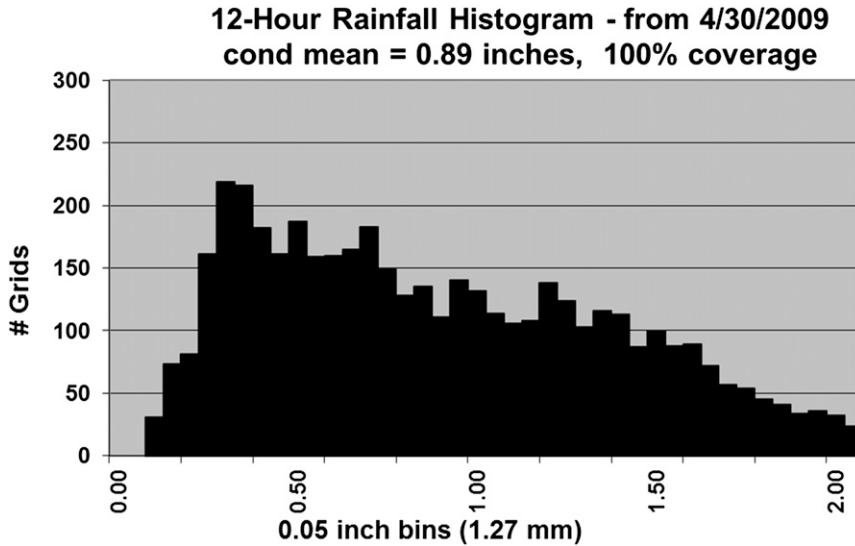


FIG. 11. As in Fig. 10, but for the 12-h period ending at 0000 UTC 30 Apr 2009.

3. Gamma distribution analyses

Analyses of the individual 12-h datasets compared each observed areal distribution to an expected gamma distribution. In the comparisons, the 12-h observed areal precipitation mean for the event was used to compute the gamma approximation for that event. Only two gamma distributions were used. Where there was 100% coverage in the 12-h period, the alpha parameter was set to 3 for the gamma PDF. For all events less than 100%, the alpha parameter was set to 1, giving an exponential distribution. The beta term was computed using the rule

$\mu = \alpha\beta$ (Wapole and Meyers 1978), where μ is the mean precipitation for the event.

A sample event is shown in Fig. 8, with an inset histogram of the rain event. There was 100% areal coverage. The main chart shows the observed cumulative distribution with rainfall labeled along the horizontal axis. The smoother curve represents the gamma expected cumulative distribution, where the precipitation mean came from the observed data. Not all curves fit this closely.

A wide range of datasets were visually inspected along with their corresponding gamma approximations. In nearly all cases, the data and gamma approximations

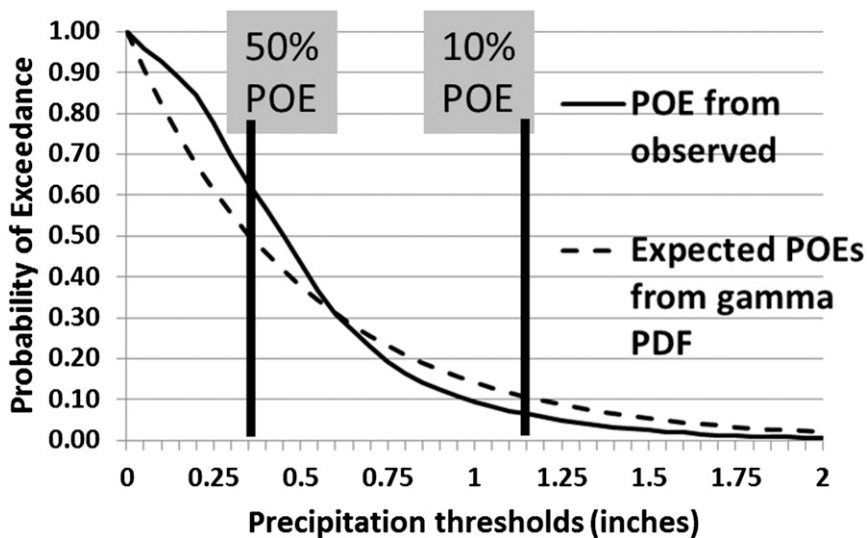


FIG. 12. POE curves computed from actual data (solid line) and from the expected gamma PDF (dashed line) using the conditional mean of the actual data. The vertical lines mark the 50% and 10% POE values. Horizontal axis is in inches.

were very reasonable. However, rigorous statistical tests, chi-squared and Kolmogorov–Smirnov, indicated that only rarely were individual events represented by gamma distributions. [McSweeney \(2007\)](#) also found that gamma approximations to observed data failed standard statistical tests. He concluded that while the fit may not be statistically significant, the approximations by the gamma distribution were good enough to represent the general form of the data. We draw the same conclusion, and further suggest that the gamma PDFs probably fit the data within the general forecast accuracy for rainfall probability and quantitative precipitation forecasts.

Other distributions, including regular, zero-inflated, and hurdle negative binomial and Poisson distributions, were used to model the binned data in an attempt to improve accuracy. However, these models did not perform as well as the gamma distribution–based models in all but a few cases. [Table 2](#) shows the various models tried and their respective mean Akaike information criteria (AIC). We note that a modified Cauchy distribution shows some promise for further investigation but is beyond the scope of this work.

a. Test details

Conditional POEs were computed from the data for each of the 12-h events where the areal coverage equaled or exceeded 10%. The 10% data cutoff was arbitrary, simply to make the dataset more manageable for computational purposes. Even after eliminating those data with less than 10% areal coverage, there were still 1800 events available for analysis.

The gamma distribution was used to compute theoretical distributions for each of the 12-h rainfall events. The observed mean conditional areal precipitation (i.e., the average rainfall from only the grids that received measurable rain) was used in each theoretical estimate. Again, the alpha term was set to 1 (exponential distribution) when the areal coverage was less than 100%, and the beta term then became the areal mean. The alpha term was set to 3 when 100% of the area received measurable rainfall and the beta term was set to the mean divided by 3, according to the rule $\mu = \alpha\beta$.

In general, an exponential PDF produced a better approximation of the data for cases where the areal coverage was less than about 95%. Where areal coverage was more than 95% and especially where areal coverage was near 100%, the gamma PDF with an alpha term of 3 produced a better approximation. Some individual cases did not always match this general rule, especially when the areal coverage was near 100%. For the purpose of this study, a more strict approach was used, as described above. [Figure 9](#) shows a comparison of the mean absolute errors for different values of alpha.

TABLE 3. The mean absolute errors computed around specific values for the probability of exceedance. Four “windows” were used except for the 10% and 90% exceedance thresholds. Again, where more than one POE value was found within a window, the POEs were averaged.

MAE around 10%	MAE	Count
POE (0.095 ≤ x ≤ 0.105)	0.0563	719
POE (0.08 ≤ x ≤ 0.12)	0.0306	1400
POE (0.05 ≤ x ≤ 0.12)	0.0287	1623
MAE around 25%	MAE	Count
POE (0.245 ≤ x ≤ 0.255)	0.0421	319
POE (0.23 ≤ x ≤ 0.27)	0.0431	975
POE (0.20 ≤ x ≤ 0.30)	0.0436	1354
POE (0.15 ≤ x ≤ 0.35)	0.0374	1557
MAE around 50%	MAE	Count
POE (0.495 ≤ x ≤ 0.505)	0.0085	159
POE (0.48 ≤ x ≤ 0.52)	0.0941	523
POE (0.45 ≤ x ≤ 0.55)	0.0970	969
POE (0.40 ≤ x ≤ 0.60)	0.0976	1251
MAE around 75%	MAE	Count
POE (0.745 ≤ x ≤ 0.755)	0.0845	84
POE (0.73 ≤ x ≤ 0.77)	0.0824	298
POE (0.70 ≤ x ≤ 0.80)	0.0833	557
POE (0.65 ≤ x ≤ 0.85)	0.0868	729
MAE around 90%	MAE	Count
POE (0.895 ≤ x ≤ 0.905)	0.0692	51
POE (0.88 ≤ x ≤ 0.92)	0.0574	144
POE (0.85 ≤ x ≤ 0.95)	0.0593	250

A crossover from alpha = 3 to alpha = 1 occurs between 95% and 100% areal coverage.

Finally, an assumption was required in computing the weighted mean rainfall for each event. One might normally select the midpoint of a data bin to compute the mean. However, since the predominant character of precipitation data is skewed toward lower rainfall amounts, it was decided to bias the computations to the lower portion of each bin. For example, the grid count for the 0.0–0.05-in. bin was multiplied by 0.02 in., not 0.025 in. Similarly, the grid count in the 0.55–0.60-in. bin was multiplied by 0.57 in., not 0.575 in. The mean for the event was then computed from the sums of the bin values. The areal coverage was a simple computation, where the number of bins with rain was divided by the total number of bins (4578).

Two sample histograms for a typical 12-h precipitation event are shown in [Figs. 10](#) and [11](#). The data show the numbers of grids in the associated precipitation bins. In [Fig. 10](#), there were just over 1000 grids that measured from 0.01 through 0.05 in. of rain and about 1100 grids that received rainfall amounts from 0.06 through 0.10 in. The shape of this histogram was quite typical for rainfall events where there was less than about 95% coverage.

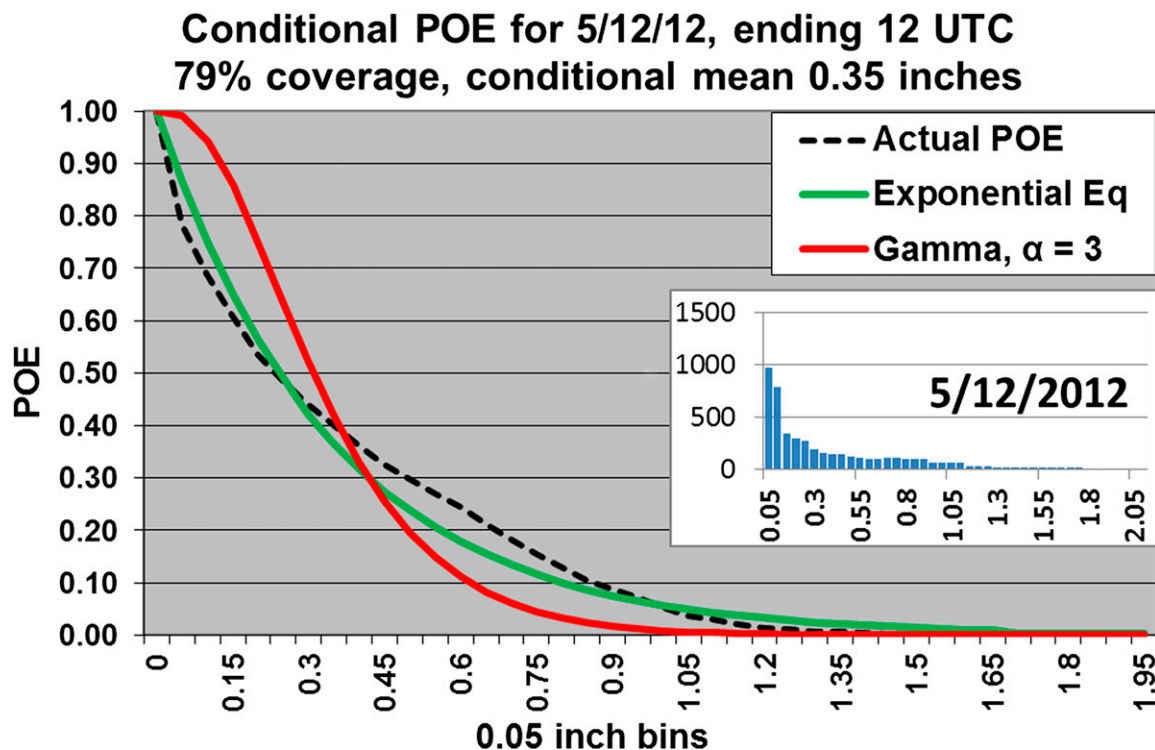


FIG. 13. Conditional probability of exceedance for the 12-h period ending at 1200 UTC 12 May 2012. Areal coverage was 79%. Conditional mean rainfall was 0.35 in.; unconditional mean rainfall was 0.27 in.

Figure 11 is typical of heavy rainfall events with near 100% coverage. The entire histogram is shifted toward higher rainfall amounts. With so much coverage, the overlap of precipitation elements over the 12-h period effectively eliminated amounts less than 0.10 in. for this particular event.

b. Test of mean absolute errors of POE

A goodness-of-fit test was conducted using MAEs between the POEs computed from the data and the expected POEs derived from the integrated gamma PDFs. Five thresholds from the observed data were used for the POEs: 90%, 75%, 50%, 25%, and 10%. The areal mean precipitation values from each event were used to create the gamma PDFs.

Graphical examples of errors are shown in Fig. 12, for POEs of 50% and 10%. By inspection, we see there was a 50% chance of exceeding 0.37 in. of rainfall; that is, 50% of the grids received more than 0.37 in. and 50% received less. The gamma estimate to exceed 0.37 in. was about 60%, an error of 10%. Also from Fig. 12, it can be seen that 7% of the observed grids received more than 1.16 in. The gamma estimate to exceed 1.16 in. was about 10%, yielding an error of 3%.

POEs were computed from all the observed data at increments of 0.05 in. However, because of the frequently

rapid drop in counts from one bin to the next, it was not always possible to compute differences between the observed POEs and the gamma-estimated POEs at the threshold values. At times, observed POE values changed more than 10% simply by moving from one 0.05-in. bin to the next.

To address this issue, the MAEs between the POEs from the data and the expected POEs from the gamma PDFs were obtained for a variety of windows, as shown in Table 3. A narrow window resulted in fewer events to test because the POEs from the actual dataset may not have had a computed value in the window. As the window was expanded, the likelihood of a data POE falling in the window increased. In fact, larger windows often resulted in several data POEs falling in the window. Where there was more than one data POE, the average error was computed from all the individual errors in the window. An example of the process may be revealing.

Suppose the binned data had POE values at 0.27 and 0.21 in. For a window around 0.25, between 0.24 and 0.26, there is no available comparison. If the window for comparison is expanded from 0.23 to 0.27, the data POE of 0.27 can be compared to the gamma approximation and an MAE (using only one value) is computed. If the window for comparison is expanded again, from 0.20 to 0.30, then two comparisons are available, and the gamma

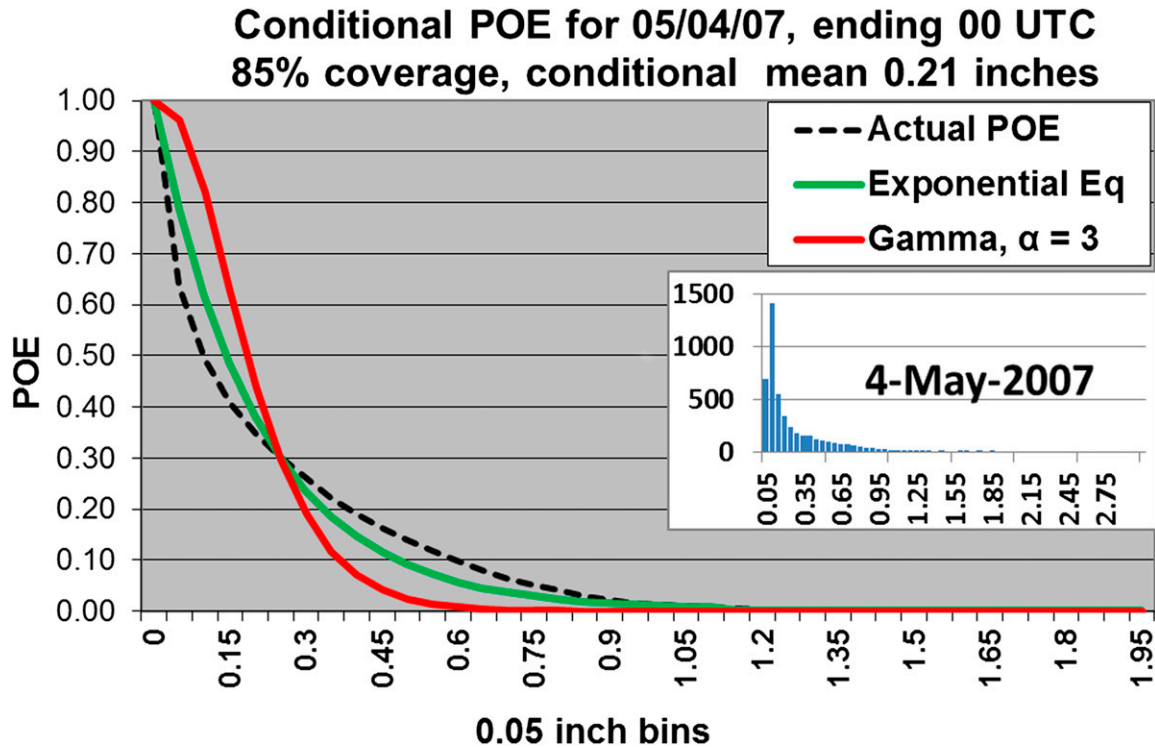


FIG. 14. As in Fig. 13, but for period ending at 0000 UTC 4 May 2007. Areal coverage was 85%. Conditional mean rainfall was 0.21 in.; unconditional mean rainfall was 0.18 in.

differences at 0.27 and 0.21 are averaged for the MAE for the window. This method was used to compute the MAEs in Table 3. Obviously, as the window for comparisons increased, the counts for comparisons also increased.

At times, especially where the areal coverage was low or where there was not a large variation in precipitation amounts, the slope of the POE curves was often quite steep (POEs higher than about 50%). When this occurred, the magnitude of the MAE could be quite large. Below about 50% POE, the curves were not as steep and the MAEs were generally not as large. Figure 12 is a reasonable example of this characteristic.

4. Application to point probabilistic quantitative precipitation forecasts

Based on this study, it is apparent that given the areal coverage and mean areal precipitation, reasonable estimates for exceedance probabilities across an area are possible. However, National Weather Service precipitation probability forecasts are for points, not areas. To be consistent, POEs need to also represent points. Some assumptions must be accepted if this areal study is to be applied to individual points.

First, one must agree that for each event, the total precipitation received at one point could just as easily

have been received at any other point, and vice versa. This is not difficult when one considers the randomness in which precipitation patterns occur. For most events, rainfall footprints could easily have been displaced several kilometers one way or the other for essentially those same meteorological variables. Over an infinite number of similar events, each point should receive the same distribution of precipitation amounts, as suggested by the law of large numbers. Hughes (1980) suggests that over a reasonably homogeneous area, the probability of occurrence at a single gauge over time will eventually match the probability of occurrence for all gauges in the area, averaged together for a single event.

Second, one must agree that the areal coverage of each event in the study is representative of the probability that any point from the area would have received rain. Hughes also suggested that a perfect forecast of the average point probability would be that of the observed areal coverage. Consider that for an infinite number of similar events, the probability of rainfall at any point will eventually match the areal coverage where the meteorological parameters are reasonably similar.

Of additional importance for National Weather Service purposes, PoPs are point forecasts. Therefore, to

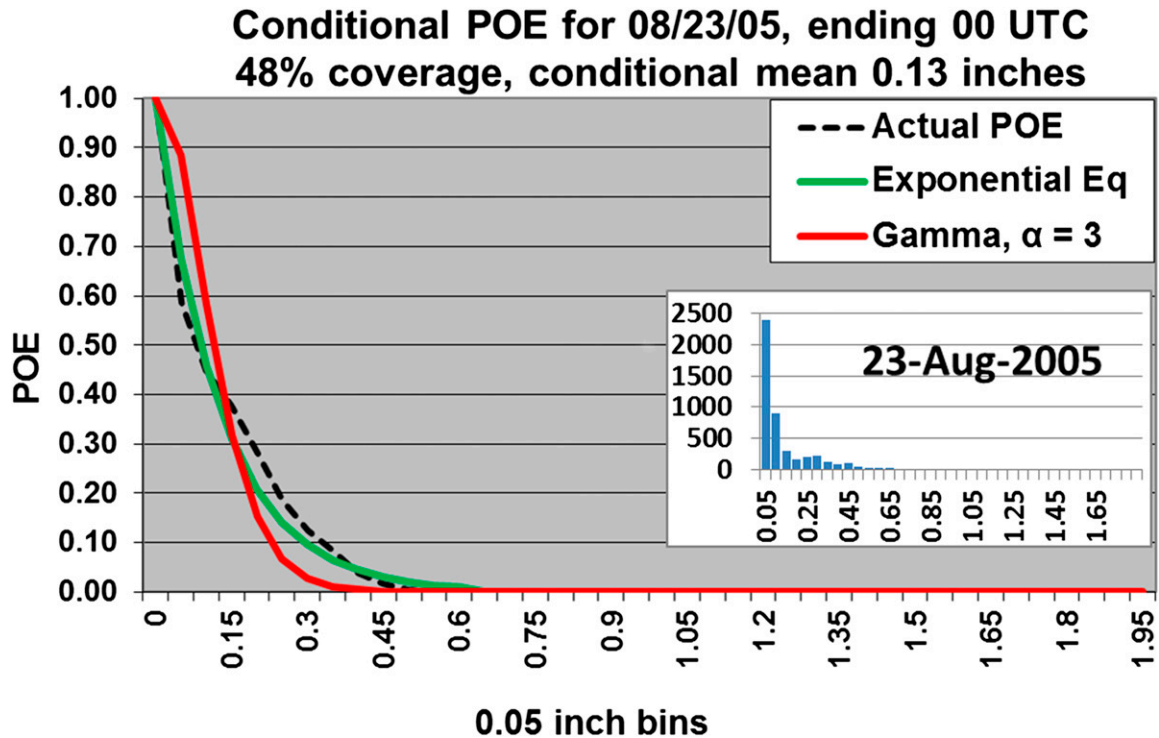


FIG. 15. As in Fig. 13, but for period ending at 0000 UTC 23 Aug 2005. Areal coverage was 48%. Conditional mean rainfall was 0.13 in.; unconditional mean rainfall was 0.06 in.

remain consistent, NWS POEs should also be for points. The data for this study come from individual events across a 73 000 km² area. However, it is not difficult to see that the areal coverage and areal precipitation could be downsized to smaller and smaller areas where meteorological parameters are more uniform. That downsizing results in modified precipitation analysis and modified areal coverage to the limit of each point. Therefore, when a forecast is made, the probability of rain at a point can be combined with the mean of the expected distribution of rainfall at that point. From the expected distribution, the unconditional probabilities of exceedance for any rainfall amount can be calculated using either of the following two equations:

$$uPOE(x) = PoP \times e^{-x/\mu}, \tag{1}$$

where μ is the mean conditional quantitative precipitation forecast (cQPF) and x is the exceedance threshold, or

$$uPOE(x) = PoP \left(\frac{x^2}{2\beta^2} + \frac{x}{\beta} + 1 \right) e^{-x/\beta}, \tag{2}$$

where $\beta = \mu/\alpha = (cQPF/3)$.

Equation (1) is derived from the integrated exponential distribution and was used to compute the MAEs

in Table 3 when the areal coverage (probability of rain) was less than 100%. Equation (2) is from the integrated gamma distribution where the alpha term is 3. Equation (2) was used to compute MAEs in Table 3 where the areal coverage was 100%. In Eqs. (1) and (2), the mean conditional precipitation is used to compute the “conditional” exceedance probabilities. This computation is then multiplied by the PoP (the probability that precipitation will even occur) to arrive at the unconditional probability of exceedance. This ensures the base NWS PoP is always higher than the probability of exceedance for other rainfall amounts. After all, the NWS PoP is itself a probability of exceedance: the probability to exceed zero rainfall. Also, unconditional QPF is easily converted to conditional QPF by dividing the unconditional QPF by the expected PoP.

A short example of the application may be helpful. We will assume that a point has a PoP of 60% and a mean expected point rainfall amount (QPF) of 0.70 in.; what, then, is the probability of receiving 1.5 in. of rainfall? Since the PoP is less than 95%, we will use Eq. (1). We must first compute the conditional rainfall amount by dividing the QPF by the PoP (0.70 in./0.60), yielding 1.17 in. From Eq. (1), the unconditional probability to exceed 1.5 in. then becomes

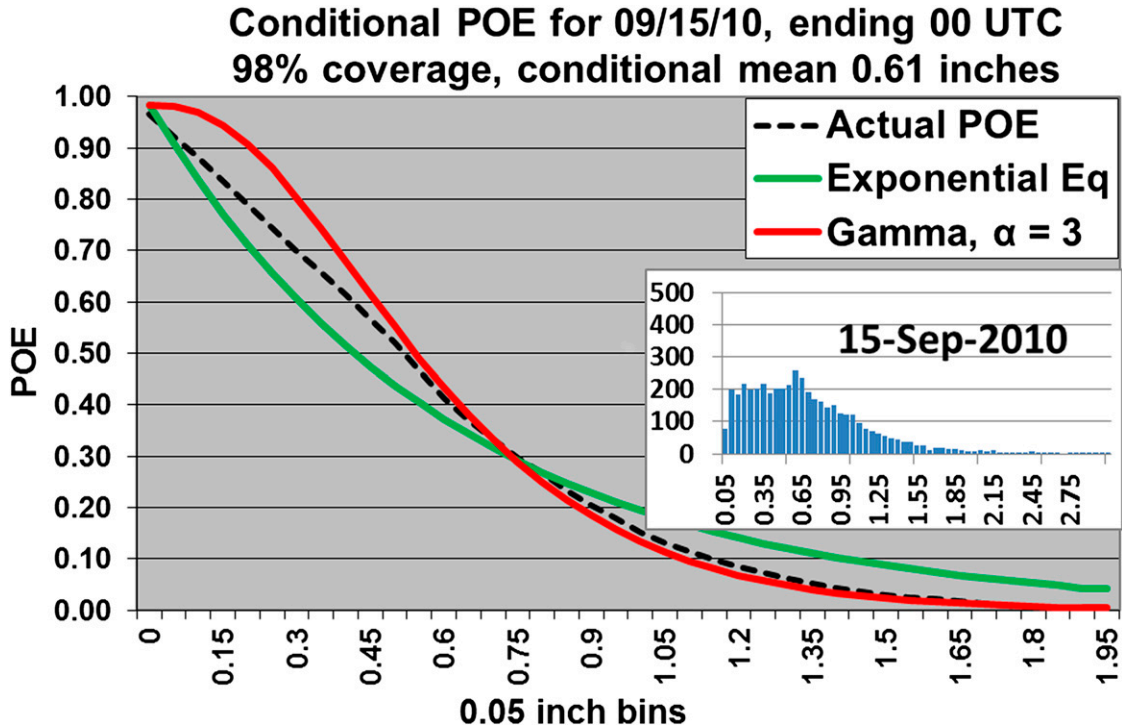


FIG. 16. As in Fig. 13, but for period ending at 0000 UTC 15 Sep 2010. Areal coverage was 98%. Conditional mean rainfall was 0.61 in.; unconditional mean rainfall was 0.60 in.

$$uPOE(x) = PoP \times e^{-x/\mu}, \tag{3}$$

$$uPOE = (0.60) \times e^{-1.5/1.17}, \text{ and} \tag{4}$$

$$uPOE = (0.60)(0.277) = 0.166 = 16.6\%. \tag{5}$$

Equation (2) can be used in a similar fashion for high-probability rainfall events. There are only two inputs: PoP and the QPF, where the QPF is the mean of the rainfall distribution a forecaster might expect over an infinite number of similar events.

Finally, a reasonable question may be posed as to whether a 4 km × 4 km grid is sufficiently small to effectively represent a point for each event. Although the general sizes of showers and thunderstorms vary considerably, a general approximation might be that a small thunderstorm could have a diameter of around 10 km (6.2 mi). This results in an area of about 78.5 km², or nearly five 16 km² grids. Furthermore, since most precipitation elements translate horizontally, the resulting precipitation footprint will sweep out a much larger area than that of a stationary precipitation element. From this, it is suggested that a 4 km × 4 km grid is a reasonable “point” approximation, at least for the purpose of this study, and within the accuracy of PoP and quantitative precipitation forecasts in general.

The analysis of these data indicated that when individual precipitation elements remained well scattered

(low areal coverage), the exponential PDF worked quite well in estimating the precipitation distributions. However, as the areal coverage approached 100%, precipitation elements began to overlap each other, reducing the areas where very low rainfall amounts occurred and increasing the areas where higher rainfall amounts occurred. In fact, many of the 100% coverage events had no grids where data fell into the 0.01–0.05-in. bin. In a few rainfall events, all grids received more than 0.15 in. of precipitation. This kind of distribution is more accurately modeled by the gamma distribution where the alpha parameter is 3.

5. Examples

Several examples (Figs. 13–19, 21) are provided here to show the reader typical gamma fits. Each example shows conditional POEs derived from the data, POEs from the exponential PDF, and POEs from the gamma (α = 3) PDF. Also included are the areal coverage and the mean areal precipitation. These figures show an inset of the histogram for each case. The dashed curves represent the probability of exceedance computed from the histogram. The green solid curve represents POEs computed from the gamma PDF where the α term equals 1 (gamma-1 exponential distribution.) As indicated earlier, this curve was determined to have a better average fit when the areal coverage was less than about 95%. The red curve shows the POEs computed from the gamma

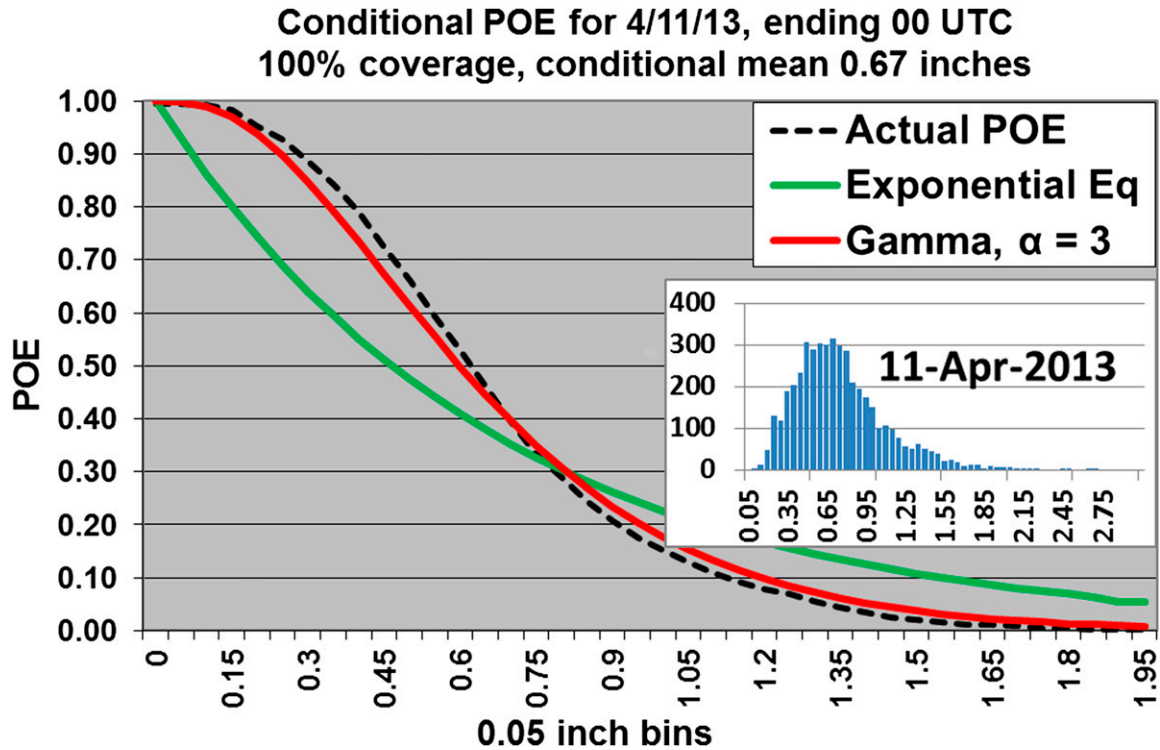


FIG. 17. As in Fig. 13, but for period ending at 0000 UTC 11 Apr 2013. Areal coverage was 100%. Conditional mean rainfall was 0.67 in.; unconditional mean rainfall was 0.67 in.

PDF where the α term equals 3 (gamma 3). This generally provided a better average fit to the observed data when the areal coverage was greater than 95%. Mean areal precipitation, areal coverage, and the dates are provided with each figure. Also, the data were collected in units of inches, so the legends were shown in inches.

Figure 13 is a convective case where the areal coverage was 79%. The conditional mean precipitation was only 0.35 in. Given the areal coverage, the gamma-1 or exponential PDF clearly had a better fit to the data. The greatest differences occurred around 0.10 in. and again around 0.60 in. One can see in the histogram that there was a slight increase in the number of grids for bins just above 0.65 in. The unconditional POE can be computed simply by multiplying the conditional POE by the areal coverage. For example, where the conditional POE is 50%, the unconditional POE is 39.5% ($50\% \times 79\%$).

Figure 14 is another convective case where the areal coverage is slightly higher at 85%. The conditional mean was lower at 0.21 in., indicating the convection did not produce rainfall as intense as in Fig. 13. Again, the exponential PDF performed better, as expected. Note that the greatest error in the model occurred between 0.10 and 0.20 in., where the computed and actual POE difference was more than 10%. This is typical of where the largest errors occur, that is, in the highest POEs where

the slope of the decline in probabilities is most steep. The error at 20% POE is only about 3% and the error at 10% POE is about 4%.

Figure 15 is a convective case in August with even lighter rainfall. In this particular case the areal coverage was 48% and the match to the exponential PDF was quite close. The gamma-3 curve did not do too poorly until the POEs were less than 20%. This was typical of how the gamma-3 PDFs erred on low areal coverage events: POEs too high for low precipitation amounts and too low for the higher precipitation amounts.

Figure 16 is an example of a heavy convective rain event. The areal coverage was 98% and the conditional mean precipitation was 0.61 in. Only 2% of the grids received no rainfall. This case supports averaging the exponential and gamma-3 PDFs when areal coverage is between 95% and 99%, although the exponential PDF was nearly a perfect match from about 40% POE and lower.

Figure 17 is another heavy convective rain event with a conditional mean of 0.67 in. and 100% areal coverage. The 100% areal coverage implies that the gamma-3 PDF should perform best, which it did. The histogram is nearly a model of the gamma-3 PDF. Clearly, not all examples of gamma-3 PDFs fit the data quite this well, but many did.

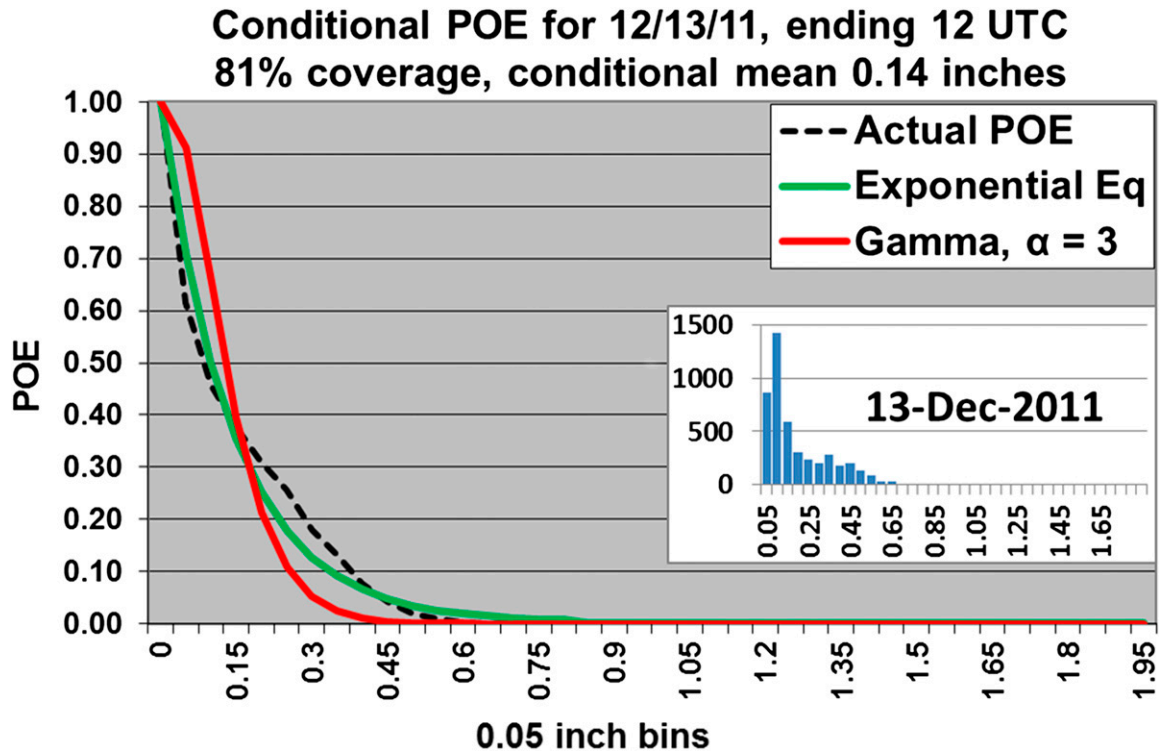


FIG. 18. As in Fig. 13, but for period ending at 1200 UTC 13 Dec 2011. Areal coverage was 81%. Conditional mean rainfall was 0.14 in.; unconditional mean rainfall was 0.11 in.

Figure 18 is a winter event with 81% coverage and only 0.14 in. of conditional mean precipitation. No convection occurred with this event. Temperatures were cool or even chilly. The exponential POEs were a nice match to the data except for some minor differences between 30% and 10% POE. Again, the gamma-3 POEs were too high for the light amounts and too low for the higher rainfall amounts, similar to the cases in Figs. 13–15.

Figure 19 is rainfall from the remnants of Tropical Storm Hermine. The storm system was convective, but had lost much of its intensity by the time it moved across eastern Oklahoma and northwestern Arkansas. The gamma-3 POE approximation was almost identical to that of the data.

Figure 20 is a radar image of the rain shield at 0600 UTC (halfway through the 12-h data accumulation period) as it moved to the east-northeast. The center of the circulation moved north-northeast.

Figure 21 was a particularly heavy rainfall event and contains a different horizontal legend than the others. Mean areal precipitation was 2.29 in. with 100% of the grids receiving rain. There was widespread flash flooding, followed by minor and moderate river and stream flooding. The heaviest rainfall totals for the 3-day event were between 6 and 9 in. The gamma-3 POEs worked well for the 12-h portion of the event shown in Fig. 21. Of

significance is that the gamma-3 POEs indicated a 10% chance of more than 4 in. during this 12-h period. In reality, the POE for 4 in., computed from the data, was only about 5%. But, from the perspective of an emergency manager or city manager, the slightly overstated probability of heavy rain was likely more helpful than detrimental.

6. Operational potential

Providing decision makers with more objective information on which to make those decisions is a goal of this study. After it was discovered that exceedance probabilities could be computed from PoP and QPF at point locations, test implementation of the method began. The variety of customers and how they might use the information was not completely known. However, this method of creating and issuing probabilistic QPFs finally provided a way to communicate information that forecasters had known all along; that the probability of rain and the expected rainfall amount are not related.

Initially, the target customers were emergency managers. These decision makers needed more information than “rainfall amounts of near 1 in. with locally heavier amounts.” Emergency managers wanted to know the

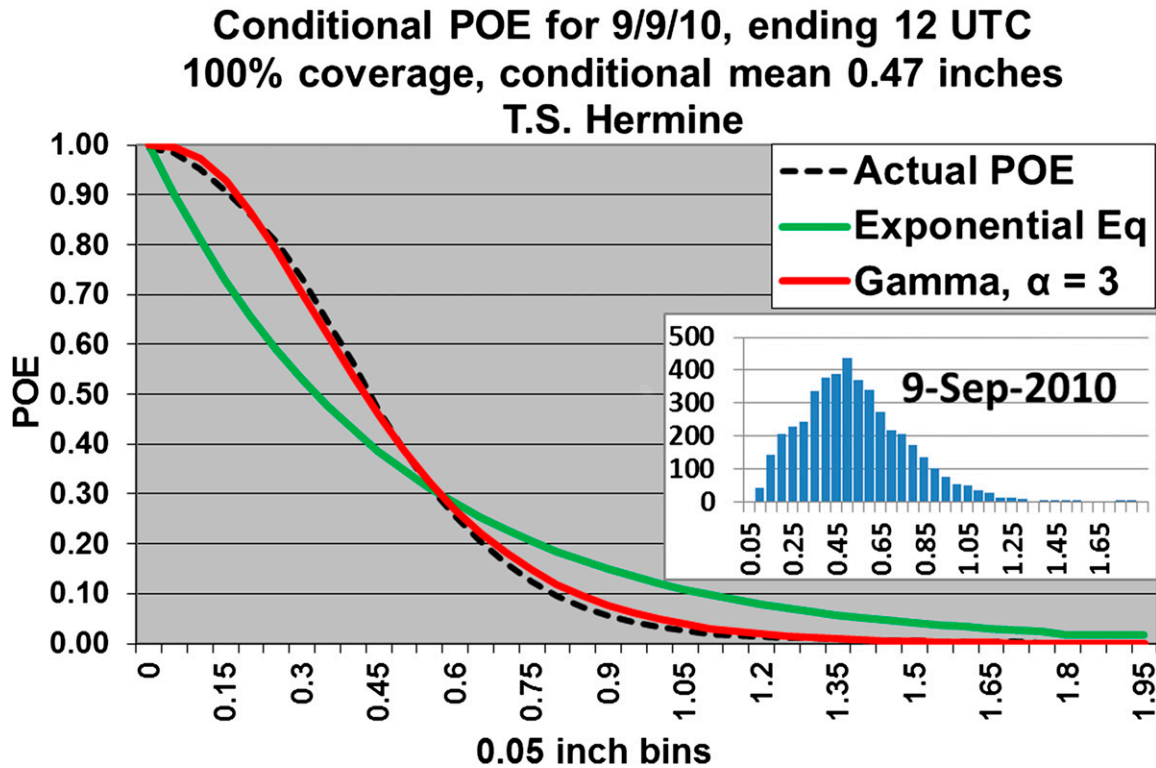


FIG. 19. As in Fig. 13, but for period ending at 1200 UTC 9 Sep 2010. Areal coverage was 100%. Conditional mean rainfall was 0.47 in.; unconditional mean rainfall was 0.47 in. This rainfall was the result of the remnants of Tropical Storm Hermine.

probability that 2 in. of rain could occur. Low water crossings might need to be barricaded. Campsites next to creeks might need to be notified. The cost–loss ratios need to be considered even for low probabilities of exceedance.

Agricultural interests may also be able to use the information in their decision-making processes. Cutting hay, irrigating fields, planting, harvesting, fertilizing, and chemical weed applications are all affected by more than the probability that 0.01 in. of rain falls. Agriculturalists want to know the probability that 0.10, 0.25, 0.50, or even 1.00 in. of rain will occur on their fields. The method suggested and confirmed here provides that information.

Also, National Weather Service Weather Forecast Offices are responsible for issuing flash flood watches. These watches are based on the probability that sufficient rainfall will result in excess runoff and rapid response flooding. The method described here can provide explicit probability estimates that those threshold rainfall amounts may occur.

An objective method of providing point forecasts for the probability of exceeding threshold rainfall amounts can be beneficial to a wide range of planners, first responders, and commercial interests. How those decision

makers use the information is different for each case, and the threshold probabilities and amounts are unique to each customer.

7. Summary and discussion

This study indicates that gamma PDFs ($\alpha = 1$ and $\alpha = 3$) provide good estimates of the distribution of precipitation over an area, given the mean precipitation and the percentage areal coverage in that area. Further, if one assumes that any point in the area could, over an infinite number of similar events, be represented by any other point in the area, then the gamma PDFs can effectively be used to provide the expected precipitation distribution at any point, given the expected mean for the point and the probability of rain at that point.

More than 7 years of 12-h precipitation data were collected over 73 248 km² at a grid resolution of 4 km × 4 km (4578 grids). These proxy rain gauges were used to create 12-h histograms for 1800 events where areal coverage was 10% and greater. From the histograms, probabilities of exceedance (POEs) were computed from the data. Those POEs were compared to exceedance probabilities derived from two gamma PDFs where the observed areal mean precipitation was used in

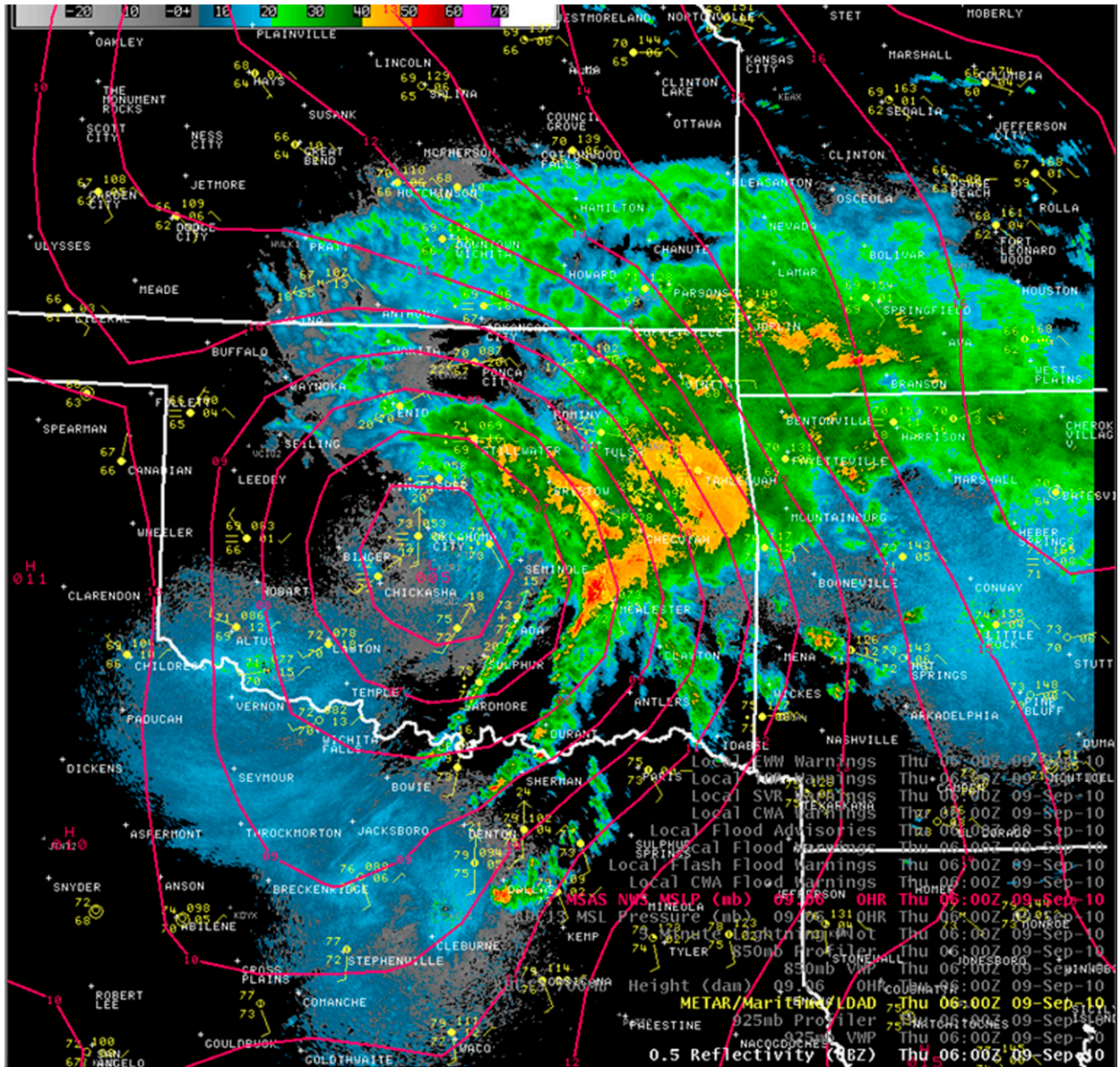


FIG. 20. Radar image of the rain shield at 0600 UTC 9 Sep 2010. The center of Hermine is near Oklahoma City, moving to the north-northeast. Rainfall is spread to the east of the center.

creating the PDFs. When areal precipitation coverage was 100%, the gamma distribution where $\alpha = 3$ was used. When the areal precipitation coverage was 99% and less, the gamma distribution where $\alpha = 1$ (exponential distribution) was used to produce the estimated exceedance probabilities. Mean absolute errors were computed at exceedance thresholds of 90%, 75%, 50%, 25%, and 10%. Results showed that the mean absolute errors from the estimates were all less than 10% of actual. The largest errors occurred at the 50% exceedance threshold. The lowest errors occurred at the 25% and 10% exceedance thresholds.

Other studies compare gamma distributions to monthly, seasonal, or annual precipitation distributions at a few points for a wide range of meteorological parameters. This study compared the gamma distribution to many points over a 12-h period with reasonably similar meteorological parameters. Therefore, this study suggests that a reasonably accurate forecast of meteorological parameters should make it possible to forecast the expected gamma distribution for the area and, within reason, any point in the area. The result is a point POE, at the scale of $4 \text{ km} \times 4 \text{ km}$, that is consistent with the National Weather Service

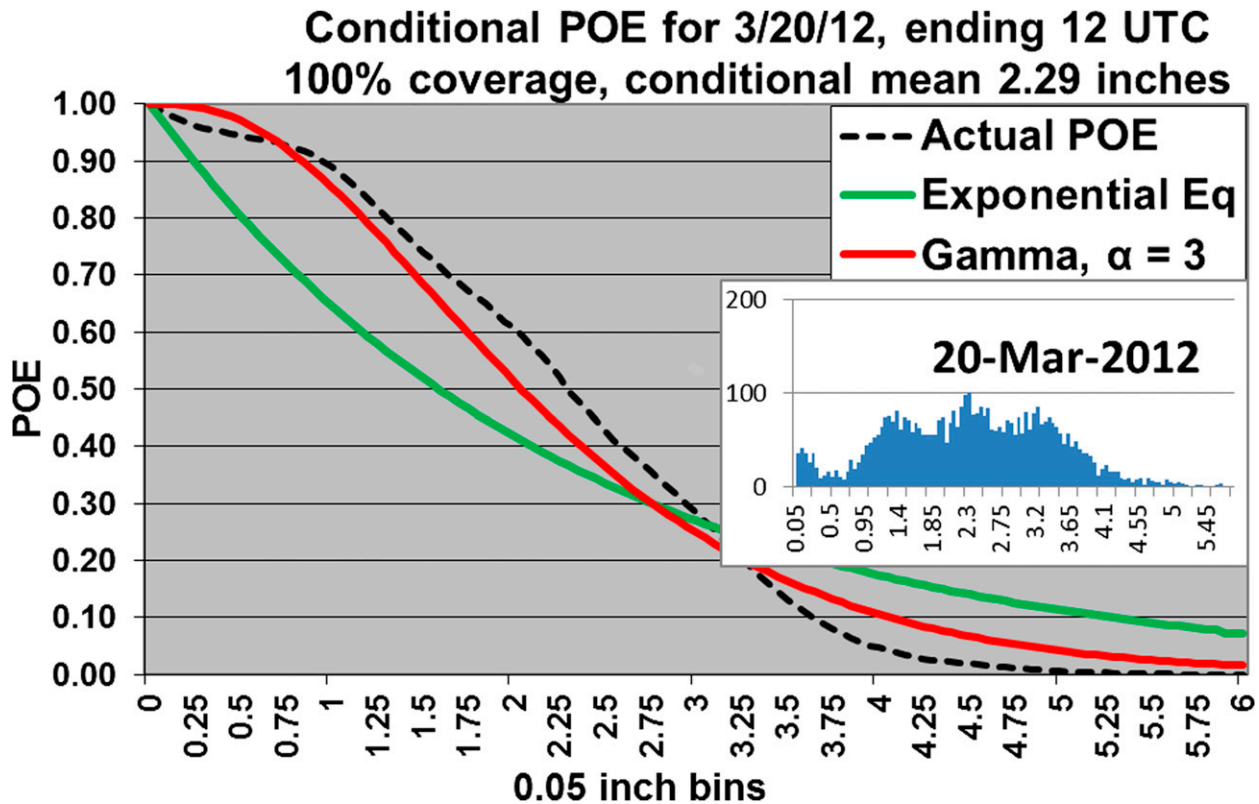


FIG. 21. As in Fig. 13, but for period ending at 1200 UTC 20 Mar 2012. Areal coverage was 100%. Conditional mean rainfall was 2.29 in.; unconditional mean rainfall was 2.29 in. This was a particularly heavy rainfall event. Please note the horizontal scale is different in this figure.

forecasts for probability of precipitation and quantitative precipitation.

The characteristic of the gamma distribution with the two different alpha terms is significant. When $\alpha = 1$, the gamma PDF (exponential PDF) diminishes immediately as rainfall amounts increase. This matches what is observed when rainfall is scattered over an area. The resulting precipitation footprints have large areas of light precipitation that decrease to near zero, while the heavier swaths of precipitation are relatively small or narrow. When areal coverage is at or very near 100%, precipitation elements generally overlap. This overlap reduces the areal coverage of light amounts and results in a PDF that does not immediately diminish as it moves away from zero rainfall. A PDF (gamma, $\alpha = 3$) holds the probabilities of exceedance near 100% for lower rainfall amounts before dropping off as the expected rainfall increases.

Furthermore, although this study used data from eastern Oklahoma and adjacent portions of northwestern Arkansas, we feel this area is simply a microcosm for any other location. The area of this study is situated at latitudes where both warm rain (collision-coalescence) and ice-crystal (Bergeron) processes are included in the data.

Point POEs have been produced at WFO Tulsa since late 2005. This POE method is now being tested by a large number of WFOs across the central and southern United States to provide decision makers with objective quantitative precipitation information that affects their areas of responsibility. The forecasts are accessible from the Hourly Weather Forecast Graph page, available from participating WFOs' websites. An explanation on how to read those forecasts can be found online as well (<http://www.youtube.com/watch?v=i2IBCSk4Mk0>).

Acknowledgments. The authors thank several individuals who helped with this study, provided insight, or otherwise provided impetus to the study. First, special thanks go to Bill Lawrence (HIC at the NWS Arkansas-Red Basin River Forecast Center, Tulsa, Oklahoma), who wrote the program to collect the data for this study. Without his help, this study would not have happened. Other thanks go to James Frederick (Information Technology Officer, NWS WFO Tulsa), who ensured the data were routinely transferred and stored for use. Dan Spaeth (NWS WFO Paducah) deserves thanks, for it was his original 1997 study on real-time PoP verification that revealed that data for this current study were

even available. For their consultations and ideas, thanks are also extended to Steven Piltz (MIC, WFO Tulsa), Nicole McGavock (Service Hydrologist, WFO Tulsa), Baxter Vieux (Vieux, Inc., Norman, OK), and Dr. Bernard Meisner (Southern Region Headquarters, National Weather Service, retired).

REFERENCES

- ABRFC, cited 2014: Arkansas–Red River Basin Forecast Center precipitation processing. NOAA/NWS. [Available online at http://www.srh.noaa.gov/abrhc/?n=pcpn_methods.]
- Hughes, L. A., 1980: Probability forecasting: Reasons, procedures, problems. NOAA Tech. Memo. NWS FCST 24, NOAA/National Weather Service, 84 pp. [NTIS PB80-164353.]
- Jorgensen, D., W. Klein, and C. Roberts, 1969: Conditional probabilities of precipitation amounts in the conterminous United States. ESSA Tech. Memo. WBTM TDL 18, 89 pp. [NTIS PB183-144.]
- McSweeney, C., 2007: Daily rainfall variability at point and areal scales: Evaluating simulations of present and future climate. Ph.D. dissertation, University of East Anglia, 256 pp. [Available online at <http://www.cru.uea.ac.uk/cru/pubs/thesis/2007-mcsweeney/4.Gamma.pdf>.]
- Spaeth, D., 1999: Alternative PoP verification at WFO Tulsa. Preprints, *15th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Dallas, TX, Amer. Meteor. Soc., 95–98.
- Thom, H., 1958: A note on the gamma distribution. *Mon. Wea. Rev.*, **86**, 117–122, doi:10.1175/1520-0493(1958)086<0117:ANOTGD>2.0.CO;2.
- Wapole, R., and R. Meyers, 1978: *Probability and Statistics for Engineers and Scientists*. Pearson, 816 pp.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.

Copyright of Weather & Forecasting is the property of American Meteorological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.