On the Emergence of Frequency Bias from Accumulating or Disaggregating Bias-Corrected Quantitative Precipitation Forecasts

BRUCE A. VEENHUIS[a] AND KEITH F. BRILL[b,a]

[a] *NOAA/NWS/NCEP/Weather Prediction Center, College Park, Maryland*
[b] *I.M. Systems Group, Inc., Rockville, Maryland*

(Manuscript received 21 May 2021, in final form 3 February 2022)

ABSTRACT: Quantitative precipitation forecast (QPF) applications often demand accumulations of precipitation for both long- and short-duration time intervals. It is desired that the shorter-duration forecasts sum to the longer-duration accumulations spanning the same time period. In the context of calibration, it is further desired that both the subinterval and longer interval accumulations be similarly corrected to have near unit frequency bias on a spatial domain. This study examines two methods of achieving these goals for 6- and 24-h accumulation intervals: 1) the accumulation method bias corrects the 6-h forecasts and accumulates them to create the 24-h accumulations; and 2) the disaggregation method bias corrects the 24-h accumulation and then proportionately disaggregates the 24-h accumulation back into 6-h accumulations. The experiments for the study are done retrospectively so that a "perfect" bias correction is possible for each method. The results of the study show that neither method accomplishes the stated goal for the calibration because QPF placement and/or timing errors contribute to frequency bias in the course of accumulation or disaggregation. However, both methods can improve the frequency bias for both the subinterval and longer interval accumulations. The choice of method may hinge most strongly on the relative tolerance of bias for the subinterval accumulations versus the longer interval accumulation.

KEYWORDS: Bias; Numerical analysis/modeling; Statistics; Forecast verification/skill; Postprocessing

1. Introduction

The weather forecast user community still demands deterministic quantitative precipitation forecasts (QPFs) even as ensemble prediction systems (EPSs) offer increasingly reliable probabilistic forecasts. Deterministic QPFs are often prepared and used over different accumulation time intervals. For example, forecasters at the National Weather Service (NWS) Weather Prediction Center (WPC) create grids of 6-h duration QPF for lead times of up to 7 days. However, the users of WPC's QPF guidance may accumulate or disaggregate the original 6-h duration QPF depending on the application. NWS Weather Forecast Offices (WFOs) may use WPC's 6-h QPF directly to prepare the National Digital Forecast Database (NDFD; Glahn and Ruth 2003) and disaggregate WPC's QPF into shorter time intervals for other products. NWS River Forecast Centers (RFC) may select a variable time accumulation interval to drive their streamflow models depending on the synoptic situation (Fread et al. 1995). Likewise government agencies such the Federal Emergency Management Agency (FEMA) may be interested in multiday QPF accumulations in the event of a landfalling tropical cyclone in order to communicate the storm total precipitation forecast. Regardless of the operation performed, it is desirable that both the original and derived accumulations represent plausible forecasts.

When preparing QPFs, forecasters often rely on post processed and statistically calibrated numerical weather prediction (NWP) output. Potential QPF calibration methods range from relatively simple approaches based on past forecast error (e.g., Beck et al. 2016; Mass et al. 2008) to much more complex methods. In their introduction, Bakhshaii and Stull (2009) list many methods of bias correction ranging from linear regression to fuzzy logic before discussing gene-expression programming, the one selected for their work.

Quantile mapping (QM; a.k.a. frequency matching) is a relatively simple and effective calibration approach now commonly used in both regional climate modeling (e.g., Reiter et al. 2018; Chen et al. 2013) and short- and medium-range weather forecasting (e.g., Hamill et al. 2017; Zhu and Luo 2015). With QM, forecast values are modified so that the frequency distribution of the forecasts matches that of the observations in the temporal domain. When calibrating NWP-based QPF, the QM method uses a historical training dataset of matched QPF and quantitative precipitation estimate (QPE) values at grid points to correct a current QPF in the manner described concisely by Hamill et al. (2017, p. 3443). In practice, QM may use either parametric or empirical (nonparametric) cumulative distribution functions (Reiter et al. 2018) to map between the predicted and observed distributions.

Whatever the method of calibration, it is expected that the bias of the corrected QPF (as measured by frequency bias, the ratio of forecast event frequency to observed event frequency) is improved compared to the raw QPF over a given verification region. Since QPFs are required for different accumulation intervals, there is a desire to have bias-corrected subinterval QPFs that sum to a bias-corrected QPF over a longer interval, such as 6-h accumulations summed to 24-h accumulations. This work examines two different methods that may be used to accomplish this goal for 6- and 24-h accumulations, and shows how each method falls short of achieving the objective due to the effects of placement and timing errors. Both methods are examined using an experimental

Corresponding author: Bruce A. Veenhuis, bruce.veenhuis@noaa.gov

DOI: 10.1175/WAF-D-21-0083.1

For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

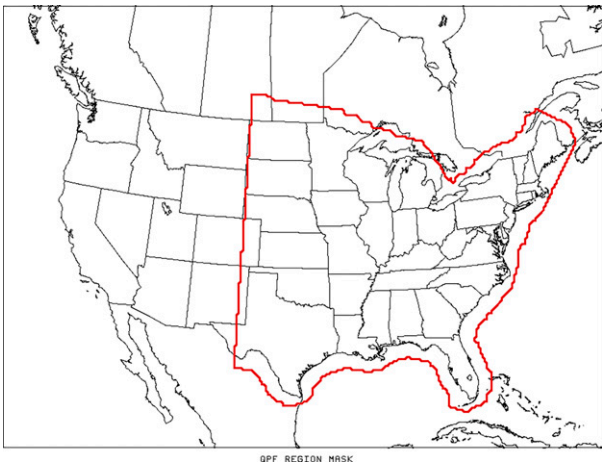

FIG. 1. The CONUS grid area is shown inside the rectangular area. The irregularly shaped region outlined in red is the domain used in the experiments.

approach applied retrospectively to deterministic QPFs and EPS mean QPFs. In the experiments, both the forecasts and the verifying analyses are available, so that the frequency bias can be adjusted to unity over a specific spatial domain. The spatial domain quantile mapping (SDQM) approach applied here is described by Clark et al. (2009, p. 1134) and demonstrated in a simple way by Pyle and Brill (2019) and in the appendix. SDQM is a nonparametric application of QM, whereby the order statistics[1] for the quantile mapping are the values at grid points covering the spatial domain. SDQM changes the QPF values so that their frequency distribution on the spatial domain exactly matches the frequency distribution of the verifying QPE analysis values. While such a perfect bias correction is certainly not possible in any real-time forecasting process, doing so in the experiments eliminates any question as to whether a better result would be obtained by improving the bias correction in some way. Although the SDQM bias correction is "perfect" (i.e., the frequency bias becomes unity), it does not alter the QPF placement and therefore is recommended in QPF verification to eliminate the influence of bias on performance metrics for QPF exceeding thresholds (Pyle and Brill 2019).

Two experiments, the accumulation experiment and the disaggregation experiment, are described in this article. For the accumulation experiment, 6-h QPFs are bias corrected on the domain outlined in Fig. 1 and accumulated to create 24-h QPFs that are verified on the same domain to ascertain the extent to which the original perfect unit bias is retained. In the disaggregation experiment, 24-h QPFs are bias corrected and then disaggregated using the constituent raw 6-h QPFs to obtain new 6-h QPFs that are verified to determine the extent to which the perfect bias is preserved for the 6-h disaggregated QPFs. Both experiments use the SDQM method of bias removal. The results of these experiments are expected to be of interest to operational forecasters who need to be aware of how biases may creep into post processed QPFs, and model post processor developers who must be mindful of the same considerations. Researchers who make use of bias-corrected simulated QPFs may find the information herein useful. These results may also encourage the research community to investigate more advanced statistical methodologies for post processing model QPFs in the day 1–7 timeframe, taking inspiration and knowledge from the extensive work already done by the hydroclimatic community (e.g., Papalexiou et al. 2018; Poschlod et al. 2018). To the authors' knowledge, this is the first published study to explore of the effects of accumulating and disaggregating bias-corrected QPF.

The remainder of the paper is organized as follows. Section 2 describes the source of deterministic and EPS mean QPFs and the QPE analyses used for the bias correction and verification. Section 3 discusses the accumulation experiment in detail and presents the results. Section 4 describes and presents results for the disaggregation experiment. A summary and conclusions are given in section 5.

2. Data sources and verification methodology

The European Centre for Medium-Range Weather Forecasts (ECMWF; https://www.ecmwf.int/en/forecasts/documentation-and-support) provides the deterministic and EPS mean (ECENS) QPFs for this work. By international agreement, the ECMWF deterministic model forecasts are available to the U.S. National Weather Service (NWS) National Centers for Environmental Prediction (NCEP) on a global grid with quarter-degree grid spacing. Similarly, the ECMWF ensemble forecasts are available on a global grid with 0.5° grid spacing. WPC maintains a long-term archive of the 6-h QPF data from the ECMWF deterministic model and the ECENS that have been remapped to a grid with 20-km grid spacing covering the conterminous United States (CONUS) on a Lambert conformal projection (see Fig. 1). To leverage this existing dataset, we used the 20-km remapped data for this study. The remapping approximately preserves area averages and is built into the GEMPAK software (https://www.unidata.ucar.edu/software/gempak/package_information/) that is used for the data processing and graphics in this work. The forecast projection times and valid date–time ranges used are given in the discussion of each experiment.

The CONUS QPE analysis is the Stage IV analysis described by Hou et al. (2014, p. 2544). The 4-km discretized QPE analysis data are remapped to a CONUS grid with 20-km grid spacing using the GEMPAK software. The Stage IV QPE covers the study domain outlined in Fig. 1. The QPE analysis is utilized for both the bias correction and the verification.

The verification applied in this study involves populating 2×2 contingency tables for dichotomous (yes/no) forecasts as described by Wilks (2006, section 7.2.1). Table 1 along with its caption describes the 2×2 contingency table. The 2×2 contingency table values may be accumulated over time and space. In terms of values in Table 1, frequency bias is $(a + b)/(a + c)$. In the case of perfect unit frequency bias,

[1] Order statistics refers to data values that have been sorted from lowest to highest value.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

TABLE 1. Typical 2 × 2 contingency table of accumulated counts or frequencies (a, b, c, d) for observed and forecast accumulated precipitation exceeding or not exceeding some specified threshold Q.

Outcomes	Observation ≥ Q (yes)	Observation < Q (no)
Forecast ≥ Q (yes)	a	b
Forecast < Q (no)	c	d

$b = c$ in Table 1, and, when such tables are combined, unit frequency bias is preserved. See the appendix for specific examples of populating 2 × 2 contingency tables.

Aside from frequency bias, the only performance metric used in this study is the Gilbert skill score (GSS, a.k.a. the equitable threat score; see Wilks 2006, p. 267). Assuming the values in Table 1 have been normalized by the total data count ($a + b + c + d$), then $\text{GSS} = (a - a_{\text{ref}})/(a - a_{\text{ref}} + b + c)$, where $a_{\text{ref}} = (a + b)(a + c)$ which gives the value of a expected for a random forecast. Many performance metrics can be computed using 2 × 2 contingency table values, and all are influenced by bias in such a way that worsening the bias can yield an improved value of the metric (Brill 2009). As explained and quantified by Brill (2009), regardless of the original frequency bias, if the removal of "yes" forecasts to reduce bias removes more false alarms than hits, then the GSS may increase. Similarly, if the addition of "yes" forecasts to increase bias adds more hits than false alarms, then the GSS may increase. Frequency bias alone does not measure forecast skill because it is not sensitive to hits.

For each comparison, the statistical significance of pairwise differences in frequency bias or GSS is assessed using the random resampling method described by Hamill (1999). The 95% confidence intervals for statistical significance (5% test level) are displayed as barred line segments in the graphics.

Limiting the study region to the eastern two-thirds of the CONUS (Fig. 1) is done for two reasons. First, the inclusion of the mountainous west would skew the order statistics for the SDQM due to the opposing seasonal regimes. In terms of precipitation maximums, in winter the west is relatively wetter and the east is relatively drier; in summer the opposite tends to occur. Second, the simple accumulation of 2 × 2 contingency table values over a spatial domain with too much variation in the event frequency should be avoided (Hamill and Juras 2006). There is noticeable, but tolerable variation in event frequency across the eastern two-thirds of the CONUS. Including the western CONUS would certainly worsen the effects described by Hamill and Juras (2006).

Another issue regarding QPF verification is mostly a matter of perspective: Is the QPF error best described as a timing error or a placement error? At a fixed point with the accumulation of precipitation taken as a function of time, QPF errors may appear as timing errors. At a fixed time with the accumulation of precipitation viewed over an extended area, QPF errors often may best be described as placement errors. In a realistic spatial–temporal domain, both timing and placement errors contribute to QPF error along with the error in the magnitude itself. In this article, the two terms, "timing" and "placement," are used separately, interchangeably, or sometimes together depending on what seems most appropriate in the immediate context. Generally, for the accumulation experiment it may be best to think in terms of placement error, whereas, for the disaggregation experiment, timing error may afford the better perspective. However, simple errors in magnitude may be easily mischaracterized as either placement or timing error.

3. The accumulation experiment

The accumulation experiment is designed to address the following question: If one could perfectly bias correct 6-h QPFs on a given domain, what are the characteristics of the bias for the accumulated 24-h QPFs? The experimental process is described algorithmically as the sequence of steps outlined below:

1) Gather four 6-h QPFs from an archive of past forecasts on the 20-km grid.
2) Gather the four verifying 6-h QPE analyses and remap them to the 20-km grid.
3) Apply SDQM over an eastern CONUS domain (Fig. 1) to create four perfectly bias-corrected 6-h QPFs.
4) Accumulate the raw biased 6-h QPFs to create a 24-h QPF also having bias.
5) Accumulate the perfectly bias-corrected 6-h QPFs from step 3 to get a potentially bias-corrected 24-h QPF.
6) Accumulate the 6-h QPE analyses to obtain a 24-h QPE analysis for verification.
7) Verify the two 24-h QPFs (one from the accumulation of the raw 6-h QPFs and the other from the accumulation of the perfectly bias-corrected 6-h QPFs) to create 2 × 2 contingency tables for QPF exceeding a series of thresholds.
8) Compare the frequency biases of the original raw 24-h QPFs with the biases of the potentially bias-corrected 24-h QPFs formed by accumulating the perfectly bias-corrected 6-h QPFs.

The accumulation experiment was carried out for both the ECMWF deterministic QPFs and ensemble mean QPFs accumulated over three 24-h periods ending at the 48-, 120-, and 144-h projection times from the 1200 UTC initial cycle time. The constituent 6-h QPF accumulation periods end at projection times 18, 12, 6, and 0 h before the end of the 24-h periods. With automation of the processing, the experiment ran daily for more than one year, beginning on 2 April 2019. For brevity, only samples from this dataset are presented here. To capture seasonal extremes, the verification statistics are cumulative over the 2019 warm season [June–July–August (JJA)] and the 2019/20 cold season [December–January–February (DJF)]. The fall and spring seasons (not shown) were also examined and found to exhibit patterns of behavior similar to JJA and DJF with results falling midway between the two opposing seasons. Accumulation of the 2 × 2 contingency table statistics over multiple seasons having different event frequencies is not recommended (Hamill and Juras 2006).

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

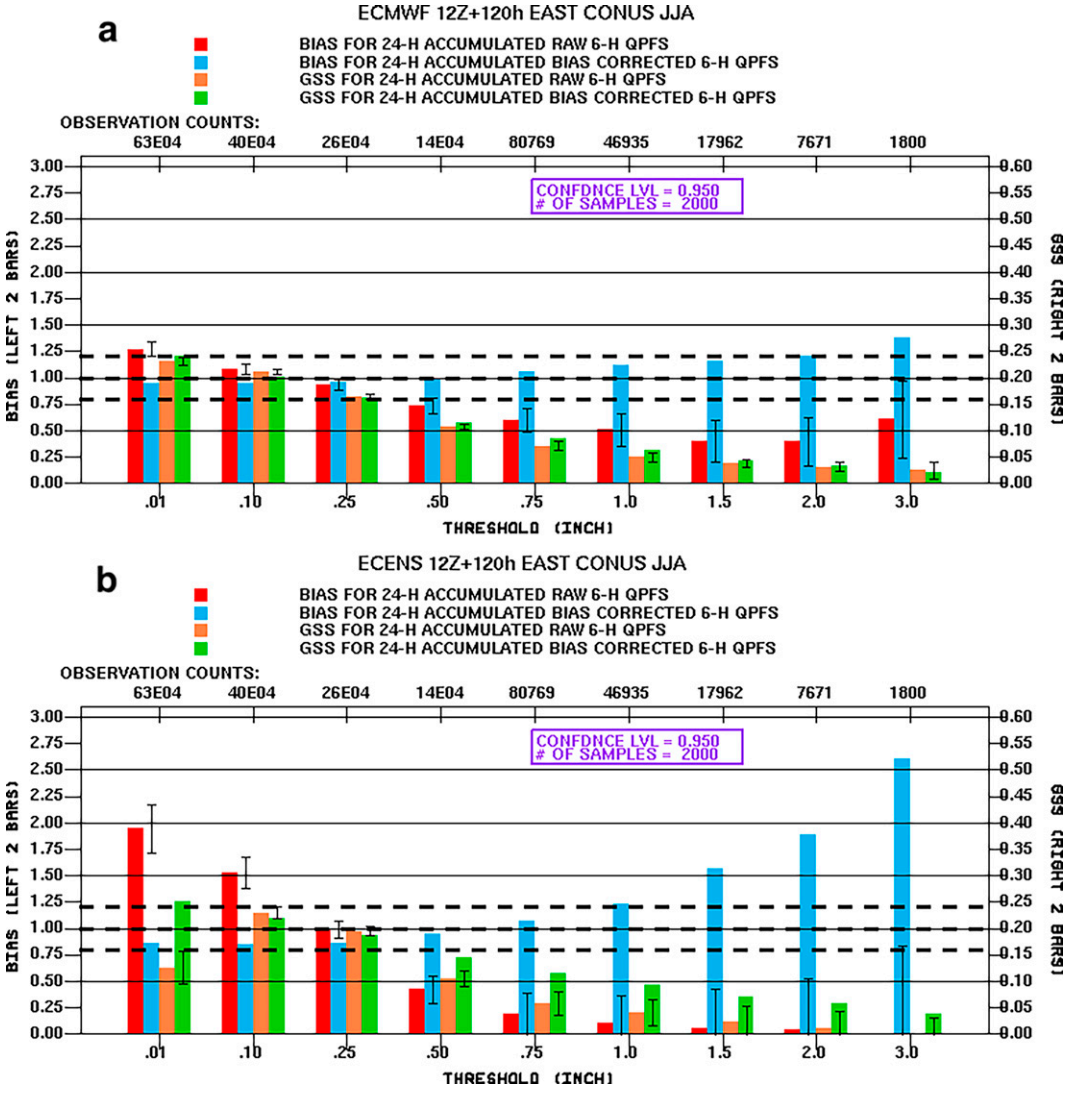

FIG. 2. JJA 2019 verification results for the 120-h 24-h QPFs from (a) the deterministic ECMWF and (b) the ECENS mean. The first pair of red and blue color-filled bars at each abscissa point compares the biases (left axis) for the raw vs bias-corrected 24-h accumulations of 6-h QPFs. The second pair of orange and green color-filled bars shows the respective GSS (right axis) comparison. Barred line segments show the 95% confidence intervals for the pairwise differences between the first and second color-filled bar of each pair. A confidence interval completely above or within a color bar indicates a statistically significant difference. Threshold values (in.) are on the lower abscissa. The values along the upper abscissa give the number of contributing grid points. Dashed bold witness lines for bias values of 0.80, 1.00, and 1.20 denote an interval of frequency bias close to unity. The inset box gives the statistical significance confidence interval and random resampling information.

Although the GSS and frequency bias generally decrease with increasing lead time, the pattern of the graphical results are qualitatively similar across projection times leading to the same conclusions; therefore, only results for the 120-h forecasts are shown.

Figure 2a shows the results for the JJA 2019 warm season for the ECMWF deterministic QPF accumulation experiment. The red and blue pairs of color bars compare 24-h QPF frequency biases for the raw versus the bias-corrected sets of 6-h QPFs accumulated to the 24-h totals that are being verified. The 24-h QPF accumulated from the raw 6-h QPFs is overbiased at the two lowest thresholds whereas the 24-h QPF accumulated from the bias-corrected QPFs is close to unity. As thresholds increase, the raw QPF underbias is improved to near unity in the midrange of thresholds. At the highest several thresholds the considerably underbiased raw QPFs are corrected to become noticeably overbiased, but only the 3.0-in. threshold exceeds 1.20. Many of the changes in frequency bias due to the correction are statistically significant at the 0.05 p level. Figure 2b

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

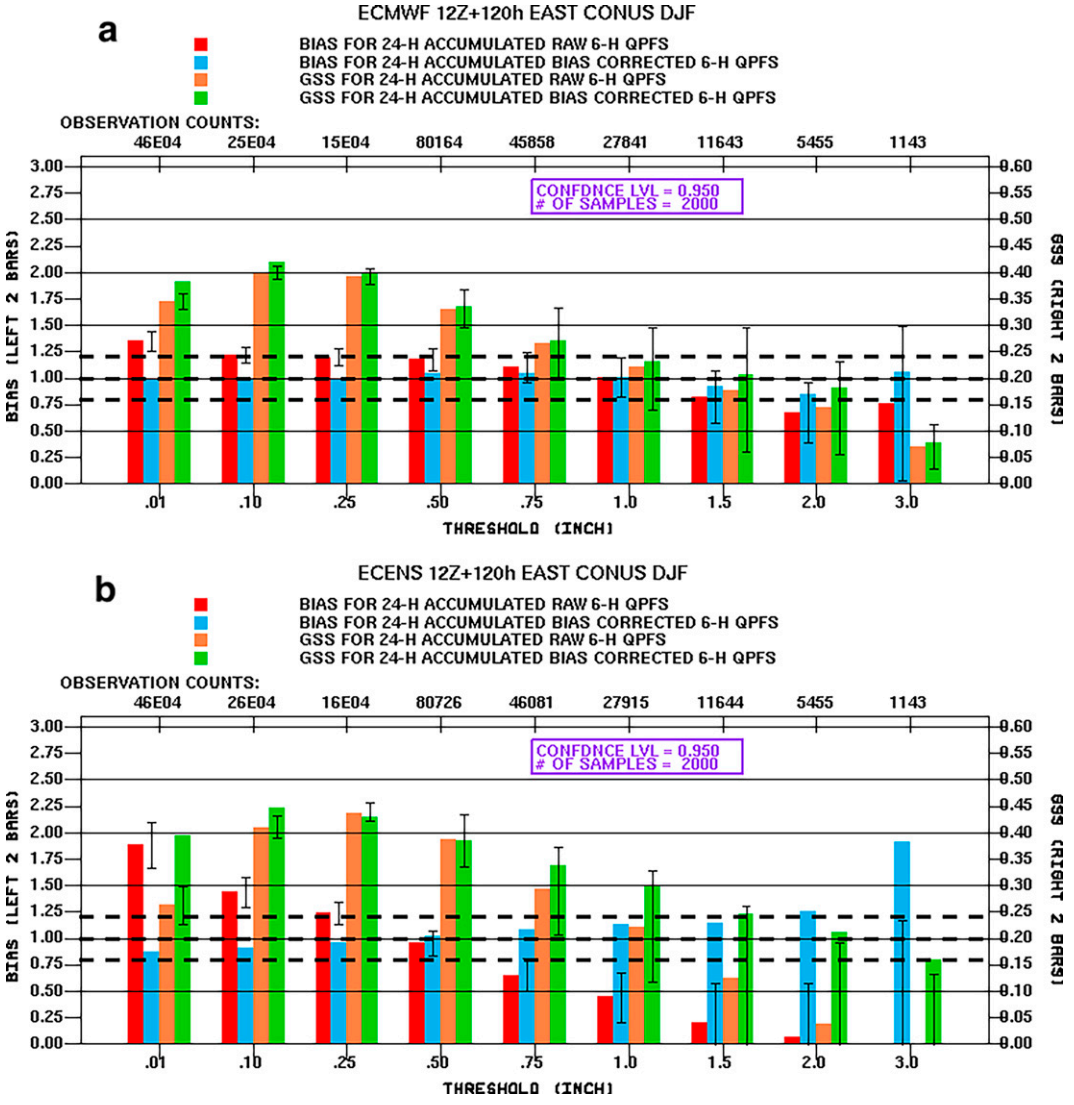

FIG. 3. As in Fig. 2, but for DJF 2019/20 verification results.

depicts results for the JJA 2019 period for the ECENS mean QPFs. The pattern of behavior for the red and blue color bars in Fig. 2b is an amplified form of that found in Fig. 2a.

In Figs. 2a and 2b, the orange and green color-filled bars compare the GSS for the raw versus bias-corrected 24-h QPFs. Except at the 0.10-, 0.25-, and 3.0-in. thresholds, the bias correction improves the GSS, and, at most thresholds, the improvements are statistically significant. The GSS is sensitive to bias and will show improved performance if the critical performance ratio condition described by Brill (2009) is met. Thus, it is expected that the GSS would increase if the frequency bias also increases. The sensitivity of the GSS to placement error and the differences in GSS between Figs. 2a and 2b provide information for understanding the amplification of the bias pattern seen in Fig. 2b compared to Fig. 2a.

Before delving into a deeper explanation of the results in Fig. 2, it is interesting to look at verification results for the DJF 2019/20 cold season shown in Fig. 3. For the deterministic QPFs (Fig. 3a), the frequency bias of the raw forecasts is within the 0.8–1.2 range except for the underbiased two highest thresholds and the overbiased 0.01-in. threshold. The bias correction produces statistically significant changes only at the lowest four thresholds for which the raw QPFs are overbiased. After the bias correction, all thresholds have frequency bias values in between 0.8 and 1.2 in Fig. 3a. Figure 3b shows the DJF results for the ECENS QPFs. Again, the general pattern of the red and blue bars in Fig. 3b is amplified compared to Fig. 3a. Overbias at low thresholds is significantly reduced, and low bias at high thresholds is significantly increased. At the highest threshold the underbias is too low to register on the graph, but the corrected QPFs are overbiased by almost a factor of 2.0 at the 3.0-in. threshold. However, it

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

is notable that the frequency biases for the corrected QPFs fall between 0.8 and 1.2 at all but the highest two thresholds in Fig. 3b.

The relatively higher GSS values in Fig. 3 compared to Fig. 2 support the well-known fact that precipitation accumulation forecasts tend to have more skill in the cold season compared to the warm season. In Fig. 3a, the bias correction leads to improved GSS at all thresholds, but the improvements are statistically significant only at the lowest two thresholds. In Fig. 3b, the GSS is significantly improved at the two lowest and the two highest thresholds. The bias correction slightly degrades the GSS at the 0.25- and 0.50-in. thresholds with the degradation at the 0.25-in. threshold being nearly statistically significant.

Figures 2 and 3 support three important points: 1) Accumulating perfectly bias-corrected 6-h QPFs to 24-h totals does not result in perfectly bias-corrected 24-h QPFs. 2) The overbias for the 24-h accumulations from bias-corrected 6-h QPFs at higher thresholds is much more pronounced for the ensemble mean than for the deterministic QPF. This effect is more noticeable in the convection dominated warm season (Fig. 2, JJA) than in the stable precipitation dominated cold season (Fig. 3, DJF). 3) Overbias at the lowest one or two thresholds is overcorrected to an underbiased condition, especially for the ECENS mean. For the last two points, less than perfect GSS values for the constituent perfectly bias-corrected 6-h QPFs (not shown) reflect the placement or timing errors that come into play to influence the biases of the accumulated 24-h QPFs.

To better understand the results described above, an idealized quantitative precipitation accumulation on a one-dimensional spatial domain (x) for four 6-h periods accumulating to a 24-h total is considered. This analysis is depicted graphically in Figs. 4 and 5. In Fig. 4, continuous functions are used to represent four 6-h accumulations on a one-dimensional spatial domain (left column of Fig. 4) that are added to obtain a continuous 24-h accumulation on the same domain (right column of Fig. 4). The raw 6-h QPFs in Fig. 4a are perfectly bias corrected using the 6-h QPEs in Fig. 4c to yield the 6-h QPFs shown in Fig. 4e. Figure 4e shows that the bias correction perfectly adjusts the accumulation amounts but cannot correct the erroneous placement of the fourth 6-h QPF maximum in the original forecast compared to the prescribed 6-h QPE in Fig. 4c. The resulting accumulation of the four perfectly bias-corrected 6-h QPFs thus produces an erroneously high-biased 24-h QPF (Fig. 4f).

Figure 5 shows two curves of frequency bias versus threshold values. The red curve in Fig. 5 results from verifying the raw 24-h QPF in Fig. 4b against the 24-h QPE shown in Fig. 4d. The blue curve in Fig. 5 results from accumulating the perfectly bias-corrected 6-h QPFs in Fig. 4e to create a 24-h total (Fig. 4f), and verifying this total against the 24-h QPE in Fig. 4d. The raw 24-h QPF accumulation is slightly overbiased at lower thresholds and underbiased at higher thresholds in Fig. 5. The opposite is true for the 24-h QPF created from the 6-h perfectly bias-corrected QPFs. Figures 4 and 5 show how placement error can lead to excessive bias at the highest thresholds for bias-corrected accumulated QPFs as is seen, especially in Fig. 2b, for actual forecasts.

To further illustrate how accumulating bias-corrected 6-h QPFs can produce an over biased 24-h QPF, a case study is shown in Fig. 6 using ECENS model data from the accumulation experiment. The results are for the 120-h forecast projection of the ECENS initialized at 1200 UTC 4 June 2020. Figure 6a plots contours enclosing areas where the raw 6-h QPF exceeds 0.5 in. for each of four sequential 6-h time periods. Figure 6b shows the 0.5-in. contours for the corresponding bias-corrected 6-h QPFs color coded as in Fig. 6a. Figure 6b also shows the 2.0-in. contours for both the 24-h accumulated 6-h bias-corrected QPFs (black) and the corresponding 24-h analyzed QPE accumulation (magenta). The 6-h raw QPFs do not accumulate in excess of 2.0 in.; hence, there is no heavy black 2.0-in. contour in Fig. 6a. In fact, it is easily seen that all four 6-h raw QPFs in Fig. 6a are under biased because the green, blue, red, and brown contours in Fig. 6b have expanded after the bias correction. In Fig. 6b, the area enclosed by the thick black contour (QPF $\geq$ 2.0 in.) is noticeably larger than the areas enclosed by the thick magenta contours (QPE $\geq$ 2.0 in.). Indeed, the computed frequency bias for the 2.0-in. threshold in Fig. 6b is 1.36; while, the frequency bias for the 2.0-in. threshold in Fig. 6a is zero. Clearly, the accumulation of the bias-corrected 6-h QPFs has led to an over biased 24-h accumulation. It should be noted, the case study presented here occurred outside of the JJA 2019 period used for the analysis in Fig. 2. However, the results for JJA 2020 (not shown) were nearly identical to those for JJA 2019.

The results for the accumulation experiment presented, discussed, and analyzed above clearly show that bias-corrected 6-h QPFs cannot be expected to sum to 24-h or other interval accumulations that will retain the same degree of bias correction. Bias correction is not preserved in the accumulation process due to the adverse influences of placement error. The latter is especially true in the case of ensemble means for which placement errors and underforecasting of amounts ultimately lead to overbias for the corrected 24-h QPF at higher thresholds. In practical terms, WPC has observed that when QM is used to bias correct 6-h ensemble mean QPFs, the accumulation of those grids may produce a significantly overbiased forecast. By significant, we mean the accumulation is higher than any individual model solution and also extreme compared to climatology. The effect is most likely to occur in the medium range (days 4–7) when there are major timing differences among the individual members in the ensemble mean. The overbias occurs because the bias-corrected 6-h ensemble mean QPFs do not progress in a synoptically realistic manner and instead repeatedly accumulate precipitation over the same geographic area. The results of the accumulation experiment motivate consideration of another method, the disaggregation method.

4. The disaggregation experiment

The disaggregation method cannot be a simple reversal of the accumulation method. Indeed, one may be well advised to

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

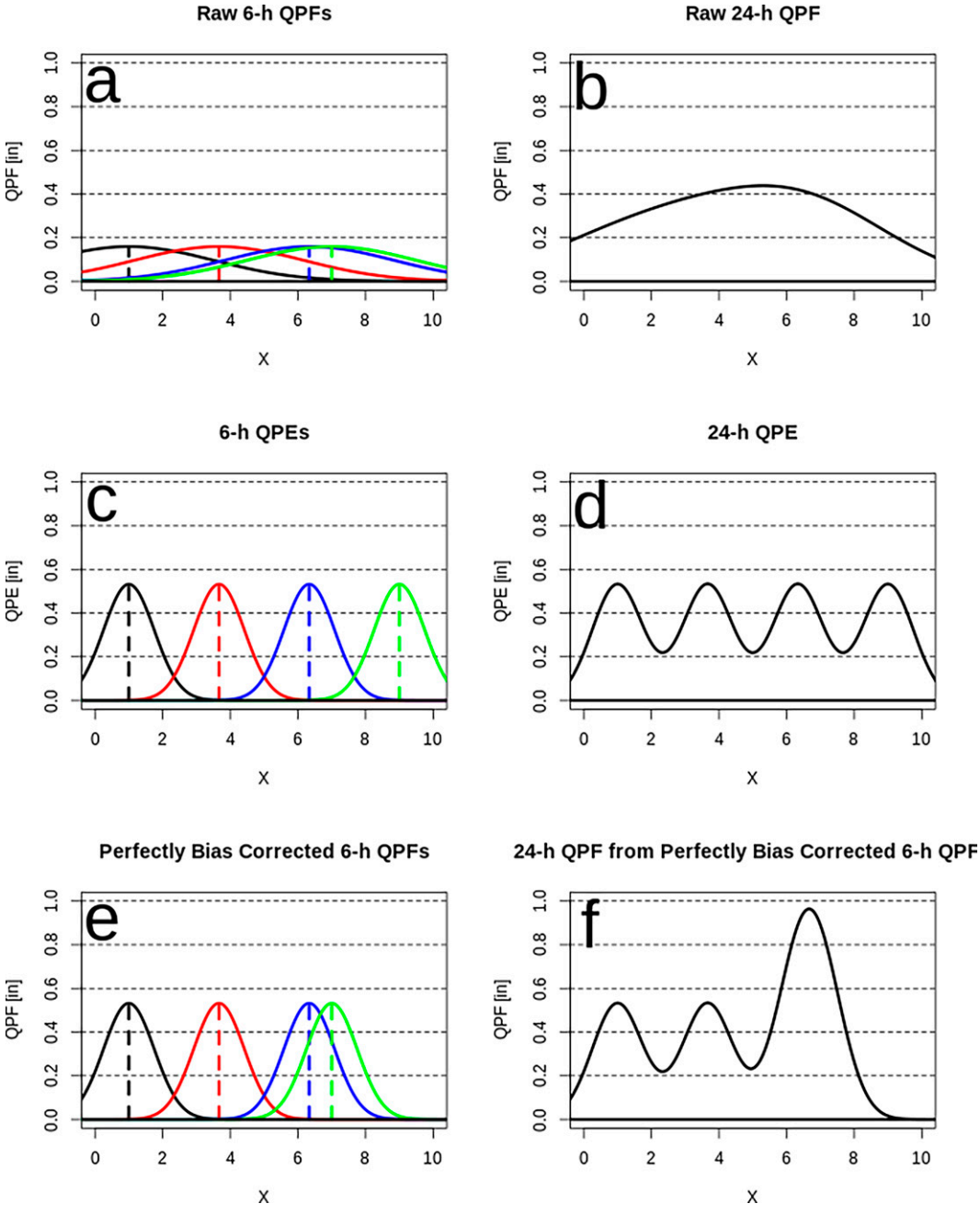

FIG. 4. (a)–(f) The 6- and 24-h continuous accumulations (ordinate; in.) as a function of position (x axis; arbitrary distance units) on a one-dimensional spatial domain. The panels are as labeled above in bold titles. The first 6-h accumulation is in black, the second in red, the third in blue, the fourth in green, in (a), (c), and (e) where dashed vertical line segments indicate placement of maximums. The 24-h accumulation is in black in (b), (d), and (f). See text for more details.

momentarily forget the preceding section to achieve the proper mindset for this discussion. The purpose of the disaggregation experiment is to assess the bias and skill of 6-h QPFs created by splitting perfectly bias-corrected 24-h accumulations into constituent 6-h accumulations. For the disaggregation experiment, the following algorithm is applied to a set of raw, biased 6-h QPFs:

1) Accumulate four biased (raw) 6-h QPFs into a 24-h QPF on the 20-km grid (Fig. 1).
2) Use the corresponding 24-h QPE to perfectly bias correct the 24-h QPF by applying SDQM on the 20-km grid.
3) Disaggregate the perfectly bias-corrected 24-h QPF into a new set of 6-h QPFs using the raw 6-h QPFs in the formulation described below.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

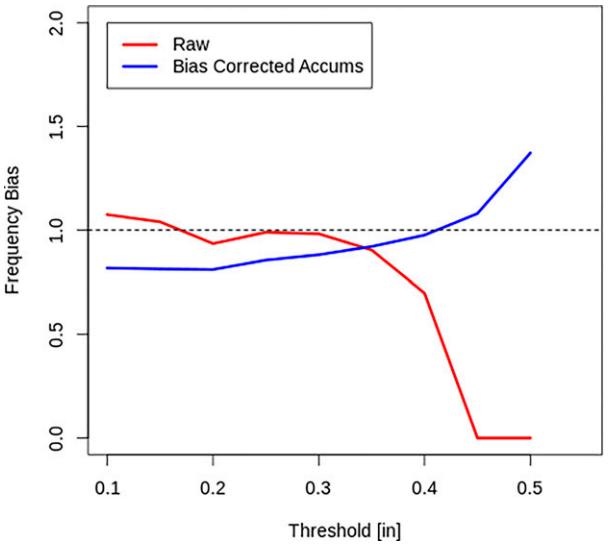

FIG. 5. Frequency bias (ordinate) as a function of threshold (abscissa; in.) for the verification of the raw 24-h QPF in Fig. 4b against the 24-h QPE in Fig. 4d (red curve) and for the verification of the bias-corrected 24-h QPF in Fig. 4f against the 24-h QPE (blue curve).

4) Verify both the original raw 6-h QPFs and the disaggregated 6-h QPFs to assess changes in frequency bias and GSS.

The disaggregation formulation applied in step 3 above redistributes the perfectly bias-corrected 24-h QPF among the contributing 6-h QPF intervals in proportion to the raw 6-h QPFs at each point on the domain according to the following equation:

$$q_{06}^{\mathrm{bc}} = q_{24}^{\mathrm{bc}}\left(\frac{q_{06}}{q_{24}}\right), \tag{1}$$

where q_{06} is a raw 6-h QPF, q_{24} is the raw 24-h accumulated QPF, q_{24}^{bc} is the perfectly bias-corrected 24-h QPF, and q_{06}^{bc} is the bias-corrected 6-h QPF corresponding to q_{06}. The disaggregation is possible only where q_{24} is greater than zero. As an example, in arbitrary units of inches, if $q_{24} = 1.00$ and $q_{06} = 0.50$, and the bias-corrected 24-h QPF is $q_{24}^{\mathrm{bc}} = 1.20$, then the disaggregated bias-corrected 6-h QPF is $q_{06}^{\mathrm{bc}} = 0.60$.

The disaggregation described by (1) was studied by Krzysztofowicz and Pomroy (1997). The quantity in parentheses on the rhs of (1) is the disaggregation fraction for any one of four 6-h subinterval accumulations within the 24-h interval. In a statistical analysis of observed precipitation data, Krzysztofowicz and Pomroy (1997) showed that given a forecast of the timing of precipitation within an interval of time, the forecast of the interval total and the subinterval disaggregation fractions can be made independently conditional on the timing. In this case, the timing of the precipitation within the 24-h interval is determined by the raw model forecasts. The bias-corrected 24-h total serves as the interval total, and the disaggregation fractions are directly forecast by the model.

The disaggregation experiment was carried out for both the ECMWF deterministic QPFs and the ECENS mean QPFs

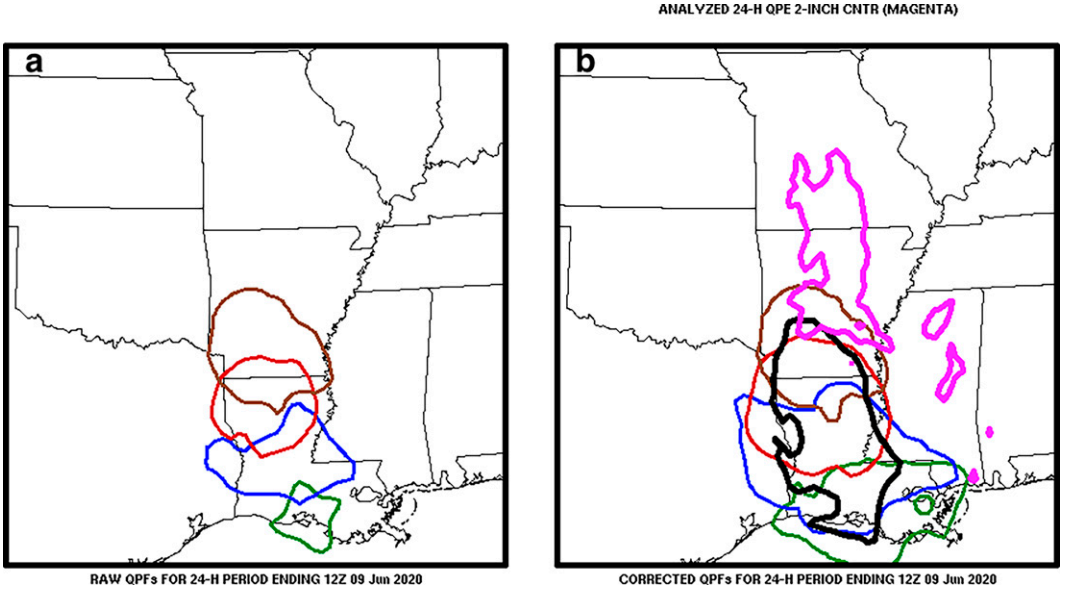

FIG. 6. Example case study for the accumulation experiment using the 120-h forecast projection from the ECENS initialized at 1200 UTC 4 Jun 2020. (a) The four raw 6-h ECENS QPFs comprising a 24-h period ending at 1200 UTC 9 Jun 2020, the valid time for the 120-h forecast. Only the 0.50-in. contours for each 6-h interval are shown. The first 6-h contour is green, the second is blue, the third is red, and the last is brown. (b) The raw 6-h QPFs are bias corrected individually, resulting in corresponding color-coded contours. The thick black contour in (b) is the 2.00-in. contour for the 24-h accumulated bias-corrected 6-h QPFs. The thick magenta contour in (b) is the verifying 2.00-in. contour for the 24-h accumulated analyzed 6-h QPEs. The raw 6-h QPFs on the left do not accumulate to 2.00 in.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

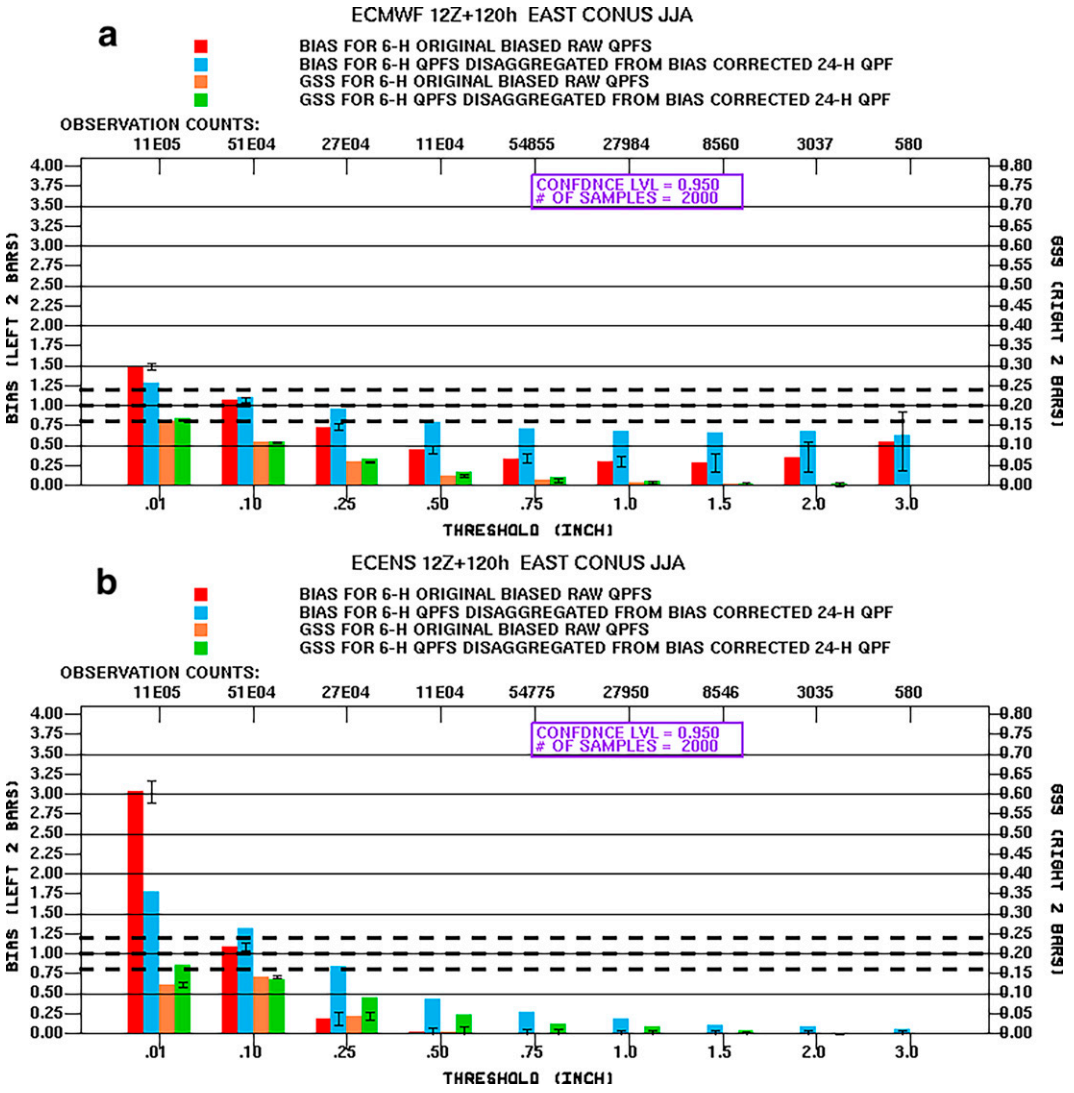

FIG. 7. JJA 2019 verification combined results for the disaggregated 6-h QPFs covering the 24-h interval ending at the 120-h projection time from (a) the deterministic ECMWF and (b) the ECENS mean. The first pair of red and blue color filled bars at each abscissa point compares the biases (left axis) for the original raw 6-h QPFs vs the 6-h QPFs disaggregated from a bias-corrected 24-h accumulation. The second pair of orange and green color-filled bars shows the respective GSS (right axis) comparison. The graphical objects and labels are as described for Fig. 2.

over the same three 24-h periods described for the accumulation experiment. These are 24-h periods ending at the 48-, 120-, and 144-h projection times from the 1200 UTC initial cycle time. Here again, results for the JJA 2019 and DJF 2019–2020 seasons are shown; not shown are results for fall and spring, which as before showed a mix of the JJA and DJF results. As with the accumulation experiment, it is sufficient to discuss in detail only the 120-h projection time since the same general pattern of the graphical results occurs for all three projection times studied. In the accumulation experiment, 6-h QPFs were bias corrected and 24-h QPFs were verified; in this disaggregation experiment, the 24-h QPF is bias corrected and the disaggregated 6-h QPFs are verified. The contingency tables for the four 6-h QPFs comprising the 24-h period are combined over the JJA or DJF verification periods in order to compute the frequency bias and GSS values.

Figure 7a shows verification results for the disaggregation method applied to the deterministic ECMWF QPF during the JJA 2019 season. The overbias at the lowest 0.01-in. threshold is corrected to a significantly lower value, but the near unit bias at the 0.10-in. threshold is increased to a slightly but statistically significant higher value. Otherwise, the underbias at most higher thresholds is corrected to statistically significant higher values, but not unit values, with the exception that the change in bias is not significant at the 3-in. threshold. At all thresholds except the lowest and highest, the GSS for the 6-h disaggregated QPFs is slightly, but significantly, improved compared to the GSS for the raw 6-h QPFs. Figure 7b shows

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

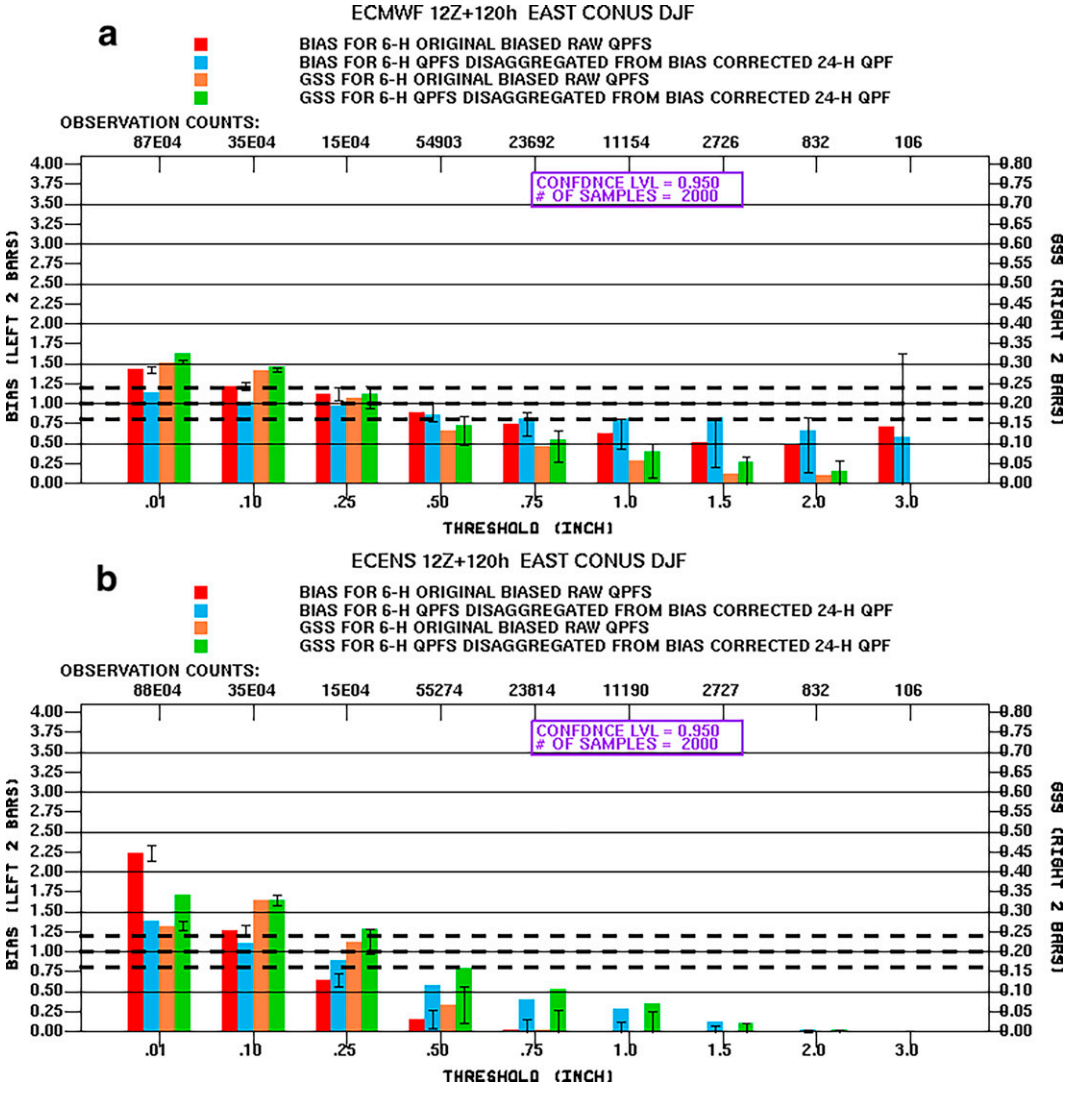

FIG. 8. As in Fig. 7, but for DJF 2019/20 verification results.

verification results for the disaggregation method applied to the ECENS mean QPF during the JJA 2019 season. As in Fig. 7a, the overbiased 0.01-in. threshold is corrected to a significantly lower, but still overbiased, value. The near unit bias at the 0.10-in. threshold is raised significantly to an overbiased value. The ECENS mean QPF is noticeably underbiased at the higher thresholds due to the smearing effect in averaging over members having different timings (and possibly placements) of the 6-h QPF amounts. These underbiases at higher thresholds are significantly improved for the disaggregated 6-h QPFs compared to the raw 6-h QPFs, but unit values are not achieved. In Fig. 7b, the disaggregated 6-h QPFs generally have significantly higher GSS values, except at the 0.10-in. threshold where both frequency bias and GSS are significantly degraded.

Verification outcomes for the DJF 2019/20 season are shown in Fig. 8. The frequency bias of the deterministic ECMWF 6-h QPFs is improved by the disaggregation method at all thresholds in Fig. 8a, except at the 0.50- and 3-in. thresholds where an underbias is made slightly, but not significantly worse. Not all of the frequency bias improvements are statistically significant, but the changes generally place the frequency bias within the 0.8 and 1.2 range. The GSS is improved at all but the highest threshold in Fig. 8a, but the GSS improvements are not statistically significant for thresholds exceeding 0.10-in. Figure 8b presents results for the disaggregation experiment applied to the ECENS mean 6-h QPFs. In Fig. 8b, all frequency bias changes are statistically significant and closer to unity, but near unit values are not obtained at thresholds exceeding 0.25 in. For thresholds up to and including 2.0 in., the GSS improved for the disaggregate 6-h QPFs compared to the original, but not all changes were statistically significant. As in Fig. 8a, the highest 3.0-in. threshold is underpopulated and not informative in this case.

Comparing the GSS values of the raw original 6-h QPFs at the thresholds above 0.25-in. between Figs. 8a and 8b, it is

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

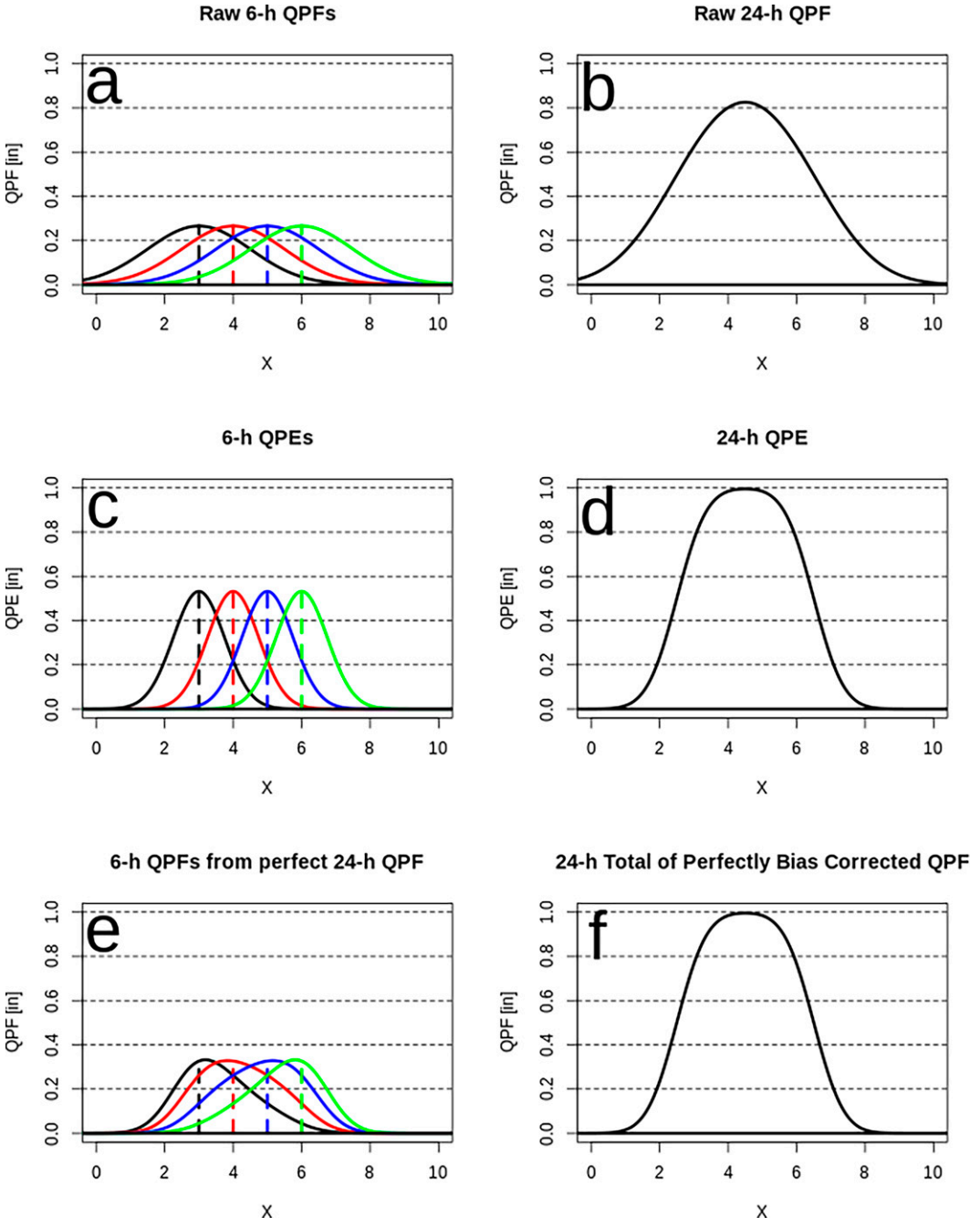

FIG. 9. As in Fig. 4, but for (e) the 6-h disaggregated QPFs and (f) the 24-h perfectly bias-corrected 24-h QPF.

reasonable to conclude that the raw ECMWF deterministic 6-h QPFs have better placement or timing of QPF compared to the ECENS mean 6-h QPFs. The same can be said of Fig. 7, but with lower GSS values. Since bias can be perfect in the presence of placement error, the better GSS values at higher thresholds in Figs. 7a and 8a indicate better timing of 6-h QPF in the deterministic ECMWF QPFs versus the ECENS mean QPFs, because the ECENS members often disagree on the timing of events by the 120-h projection time. With reference to (1), better timing of the raw 6-h QPFs (q_{06}) will give better results for the frequency bias and also the GSS of the corrected 6-h QPFs (q_{06}^{bc}). This is because (1) will place the 6-h accumulations proportionately in the correct 6-h intervals.

To provide some insight into the results described above, idealized analytic functions on a spatial domain (x) are used to represent raw 6-h QPFs and QPEs for a 24-h period, allowing the computation of disaggregated 6-h QPFs as shown in Fig. 9. The construction of these examples is similar to that

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

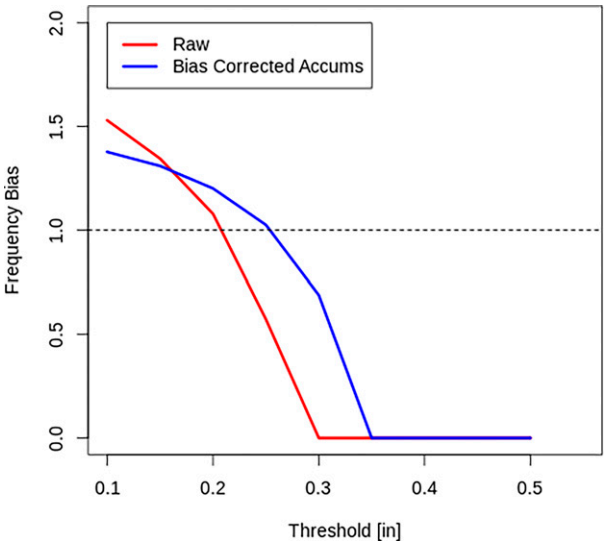

FIG. 10. Frequency bias (ordinate) as a function of threshold (abscissa; in.) for the verification of the raw 6-h QPFs in Fig. 9a against the 6-h QPEs in Fig. 9c (red curve) and for the verification of the disaggregation bias-corrected 6-h QPFs in Fig. 9e against the 6-h QPEs (blue curve).

done for the accumulation experiment (Figs. 4 and 5). The perfectly bias-corrected 24-h QPF shown in Fig. 9f is created by quantile mapping of the raw 24-h QPF in Fig. 9b to the frequency distribution of the 24-h QPE shown in Fig. 9d. This perfectly bias-corrected 24-h QPF is disaggregated using the raw 6-h QPFs shown in Fig. 9a, yielding the four disaggregated 6-h QPFs shown in Fig. 9e. Both sets of the four 6-h QPFs, the raw QPFs in Fig. 9a and the disaggregated 6-h QPFs in Fig. 9e, are verified against the corresponding 6-h QPEs in Fig. 9c. The verification involves creating 2×2 contingency tables for multiple thresholds, allowing the display of frequency bias versus threshold for both the raw and disaggregated 6-h QPFs shown in Fig. 10. The relationship of the two bias graphs in Fig. 10 is similar to what is seen for actual data in Figs. 7 and 8: overbias at the lower thresholds is corrected, but not perfectly, to lower values, whereas, underbias at the higher thresholds is corrected toward higher values, but not perfectly.

Like the accumulation method, the disaggregation method cannot produce a set of 6-h QPFs that sum to a 24-h QPF so that both the 6- and 24-h QPFs have the same degree of bias correction. Unlike the accumulation method, the disaggregation approach behaves more consistently whether it is applied to a deterministic QPF or an ensemble mean QPF in that large inflations of bias beyond unity are not typical for the disaggregated QPFs.

5. Summary and conclusions

Two methods for bias correcting QPFs to achieve some degree of bias correction for an accumulation interval and its constituent subinterval accumulations have been described and analyzed. For the accumulation method, subinterval accumulations are bias corrected with the hope that the interval accumulation is also bias corrected to the same degree. For the disaggregation method, the interval accumulation is bias corrected, and the subinterval accumulations are recomputed in proportion to the raw subinterval accumulations with the hope that the bias correction is conveyed to the subinterval accumulations. Since the retrospective experiments used to test these two methods allowed perfect bias correction on a spatial domain, it could easily be determined the degree to which either of the two methods accomplished the twin goals of consistent (i.e., perfect) bias correction for both the subinterval QPFs and the interval QPF. Neither method perfectly preserved the original bias correction but both did improve the frequency bias of the resulting forecasts compared to the original QPFs. The considerations and tradeoffs for each method are reiterated below.

TABLE A1. Synthetic raw QPF/QPE values on a 12-point spatial domain representing forecast and corresponding analyzed values.

0.07/0.11	0.16/0.21	0.39/0.51	0.54/0.68
0.19 0.95	0.35/0.81	0.63/0.52	0.75/0.28
0.24/0.18	0.28/0.25	0.48/0.58	0.52/0.44

The inflation of bias at higher thresholds that may occur when using the accumulation method (e.g., Fig. 2b) might be taken as a counterpoint to argue in favor of the disaggregation method. However, the perfect bias correction used in the experiment cannot be applied in practice, and any realistic real-time bias correction would likely be unable to boost underbias to unity at higher thresholds. Therefore, an inflation of bias at higher thresholds may be beneficial. The disaggregation method might also be indicated for forecast systems that produce well placed and timed subinterval accumulations characterized primarily by magnitude errors, but there is no clear outcome from the experiments to declare one method superior to the other. If it is required that subinterval accumulations sum exactly to the interval value and that both the subinterval and interval accumulations benefit from some degree of bias correction, then the choice of method must hinge on whether it is most important for the subintervals to have direct bias correction or the longer interval to have direct bias correction. If the subintervals are more important, then the accumulation method is chosen. If the longer accumulation interval is more important, then the disaggregation method is selected. Iterative processing, bouncing back and forth between the two algorithms, was beyond the scope of this work.

It should be noted, the hydroclimatic community has developed very advanced statistical methods of disaggregating precipitation accumulations to subintervals compared to the methodology used by this study (e.g., Bárdossy and Pegram 2016). Such methods are important to hydrological applications such as relating streamflow conditions to the spatial and temporal distribution of precipitation over a catchment. Papalexiou et al. (2018) provide a history of the hydroclimatic development of approaches to disaggregation, propose a new method, and apply their method to disaggregate monthly observed precipitation to hourly amounts and climate model

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

TABLE A2. Tabulated values demonstrating the SDQM process applied to the raw QPF and analyzed QPE values in Table A1. See text for details.

QPF	Analyzed QPE	Sorted QPF	Sorted QPE	SDQM'ed QPF	Analyzed QPE
0.07	0.11	0.07	0.11	0.11	0.11
0.16	0.21	0.16	0.18	0.18	0.21
0.39	0.51	0.19	0.21	0.51	0.51
0.54	0.68	0.24	0.25	0.68	0.68
0.19	0.95	0.28	0.28	0.21	0.95
0.35	0.81	0.35	0.44	0.44	0.81
0.63	0.52	0.39	0.51	0.81	0.52
0.75	0.28	0.48	0.52	0.95	0.28
0.24	0.18	0.52	0.58	0.25	0.18
0.28	0.25	0.54	0.68	0.28	0.25
0.48	0.58	0.63	0.81	0.52	0.58
0.52	0.44	0.75	0.95	0.58	0.44

monthly totals to daily values. Poschlod et al. (2018) compare a statistical method of disaggregation to a method using a numerical prediction model to go from daily to hourly precipitation accumulations. A detailed examination of the choice of disaggregation methodology and the effect on results was beyond the scope of this study but may be an area of future research.

Acknowledgments. The authors appreciate the encouragement of the NCEP Weather Prediction Center (WPC) management to pursue this study. WPC management approval of the use of the authors' time and WPC computational resources needed to complete the experiments is much appreciated. WPC also funded the publication of this work. The authors wish to acknowledge David Rudack who provided useful feedback regarding the manuscript and thank three anonymous reviewers for their comments and suggestions. Any opinions expressed in this work are solely those of the authors, not the funding agency.

APPENDIX

Spatial Domain Quantile Mapping (SDQM)

The perfect bias correction is carried out for each individual QPF field, a horizontal array of QPF values at regularly spaced point locations for a given valid date and time. For each such QPF field there exists a QPE field of analyzed observed values at the same points. For this demonstration, 3×4 arrays of synthetic QPF and QPE data (inch units) are created as shown in Table A1. The QPF values are under biased. The SDQM process will correct the bias as described below with reference to Table A2.

TABLE A3. Synthetic bias-corrected QPF/QPE values on a 12-point spatial domain representing forecast and corresponding analyzed values.

0.11/0.11	0.18/0.21	0.51/0.51	0.68/0.68
0.21/0.95	0.44/0.81	0.81/0.52	0.95/0.28
0.25/ 0.18	0.28/0.25	0.52/0.58	0.58/0.44

TABLE A4. A 2×2 contingency table for the verification of the raw QPFs in Table A1 for the 0.50 threshold.

Outcomes	QPE $\geq$ 0.50	QPE $<$ 0.50	Total count
QPF $\geq$ 0.50	2	2	4
QPF $<$ 0.50	4	4	8
Total count	6	6	12

The SDQM method proceeds from left to right in Table A2. Columns 1 and 2 just list the raw QPFs and the corresponding QPEs in "reading" order from upper left to lower right in Table A1. First, both the raw QPFs and the corresponding QPE values are sorted from lowest to highest values as shown in columns 3 and 4 of Table A2. Next, the mapping is done as follows: For each raw QPF value in column 1, find its position in column 3, and then populate column 5 with the corresponding QPE value from column 4. This procedure applied to each value in column 1 fills column 5 with perfectly bias-corrected QPFs. The last column in Table A2 repeats column 2 to show in juxtaposition the QPE values corresponding to the bias-corrected QPF values. The last two columns of Table A2 are used to populate the horizontal array of bias-corrected QPF values and QPE values as shown in Table A3, which is the bias-corrected version of Table A1. As an example, consider the raw QPF value 0.63 in column 1 of Table A2. In the sorted QPF column, 0.63 appears in the penultimate row next to the sorted QPE value 0.81 in column 4. Thus, 0.81 replaces 0.63 in the quantile mapped QPFs found in column 5.

Table A4 is a 2×2 contingency table (including marginal totals) verifying the raw QPFs in Table A1 against the corresponding analyzed QPE for the threshold of 0.50 in. For convenience and simplicity, here the critical success index (CSI, a.k.a. threat score) is used to assess the performance. In terms of the parameters in Table 1 of the main text, the CSI $= a/(a + b + c)$. Computed from Table A4, the frequency bias for the raw QPFs is $4/6 = 2/3$. Thus, the raw QPFs are under biased. The CSI is $2/8 = 1/4$.

If the bias-corrected QPFs in Table A3 are verified similarly, the resulting 2×2 contingency table with marginal totals is Table A5. After the SDQM bias correction, the frequency bias is unity ($6/6 = 1$) as expected. The CSI has improved to the value $4/8 = 1/2$. When multiple tables of counts are additively combined, the property of unit frequency bias is preserved if each contributing table has a unit frequency bias like Table A5.

There are two additional points of interest to note. First, the SDQM bias correction cannot do much to correct the

TABLE A5. A 2×2 contingency table for the verification of the bias-corrected QPFs in Table A3 for the 0.50 threshold.

Outcomes	QPE $\geq$ 0.50	QPE $<$ 0.50	Total count
QPF $\geq$ 0.50	4	2	6
QPF $<$ 0.50	2	4	6
Total count	6	6	12

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC

TABLE A6. A 2 × 2 contingency table for the verification of the bias-corrected QPFs for the right half subset of Table A3 for the 0.50 threshold.

Outcomes	QPE ≥ 0.50	QPE < 0.50	Total count
QPF ≥ 0.50	4	2	6
QPF < 0.50	0	0	0
Total count	4	2	6

placement error existing in the middle row of Table A1. The QPFs are heavy on the right (east) whereas the QPEs are heavy on the left (west). This condition still exists, although ameliorated somewhat, in Table A3. The second point has to do with subsetting the domain. The SDQM bias correction holds for the domain as a whole but not necessarily for a subset of the domain. For example, consider the right (eastern) half of Table A3 having six data points. The 2 × 2 contingency table for this subset is given in Table A6. The frequency bias for the subset is 6/4 = 3/2, indicating over bias. Although no subsetting was done in this work, it is important to note that subsets must be bias corrected individually.

REFERENCES

Bakhshaii, A., and R. Stull, 2009: Deterministic ensemble forecasts using gene-expression programming. *Wea. Forecasting*, **24**, 1431–1451, https://doi.org/10.1175/2009WAF2222192.1.

Bárdossy, A., and G. G. S. Pegram, 2016: Space-time conditional disaggregation of precipitation at high resolution via simulation. *Water Resour. Res.*, **52**, 920–937, https://doi.org/10.1002/2015WR018037.

Beck, J., F. Bouttier, L. Wiegand, C. Gebhardt, C. Eagle, and N. Roberts, 2016: Development and verification of two convection-allowing multi-model ensembles over western Europe. *Quart. J. Roy. Meteor. Soc.*, **142**, 2808–2826, https://doi.org/10.1002/qj.2870.

Brill, K. F., 2009: A general analytic method for assessing sensitivity to bias of performance measures for dichotomous forecasts. *Wea. Forecasting*, **24**, 307–318, https://doi.org/10.1175/2008WAF2222144.1.

Chen, J., F. P. Brissette, D. Chaumont, and M. Braun, 2013: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour. Res.*, **49**, 4187–4205, https://doi.org/10.1002/wrcr.20331.

Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, https://doi.org/10.1175/2009WAF2222222.1.

Fread, D. L., and Coauthors, 1995: Modernization in the National Weather Service River and Flood Program. *Wea. Forecasting*, **10**, 477–484, https://doi.org/10.1175/1520-0434(1995)010<0477:MITNWS>2.0.CO;2.

Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195–202, https://doi.org/10.1175/BAMS-84-2-195.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

——, and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, https://doi.org/10.1256/qj.06.25.

——, E. Engle, D. Myrick, M. Peroutka, C. Finan, and M. Scheuerer, 2017: The U.S. national blend of models for statistical postprocessing of probability of precipitation and deterministic precipitation amount. *Mon. Wea. Rev.*, **145**, 3441–3463, https://doi.org/10.1175/MWR-D-16-0331.1.

Hou, D., and Coauthors, 2014: Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of Stage IV toward CPC gauge-based analysis. *J. Hydrometeor.*, **15**, 2542–2557, https://doi.org/10.1175/JHM-D-11-0140.1.

Krzysztofowicz, R., and T. A. Pomroy, 1997: Disaggregative invariance of daily precipitation. *J. Appl. Meteor.*, **36**, 721–734, https://doi.org/10.1175/1520-0450-36.6.721.

Mass, C. F., J. Baars, G. Wedam, E. Grimit, and R. Steed, 2008: Removal of systematic model bias on a model grid. *Wea. Forecasting*, **23**, 438–459, https://doi.org/10.1175/2007WAF2006117.1.

Papalexiou, S. M., Y. Markonis, F. Lombardo, A. AghaKouchak, and E. Foufoula-Georgiou, 2018: Precise temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for stationary and nonstationary processes. *Water Resour. Res.*, **54**, 7435–7458, https://doi.org/10.1029/2018WR022726.

Poschlod, B., O. Hodnebrog, R. R. Wood, K. Alterskjaer, R. Ludwig, G. Myhre, and J. Sillmann, 2018: Comparison and evaluation of statistical rainfall disaggregation and high-resolution dynamical downscaling over complex terrain. *J. Hydrol.*, **19**, 1973–1982, https://doi.org/10.1175/JHM-D-18-0132.1.

Pyle, M. E., and K. F. Brill, 2019: A comparison of two methods for bias correcting precipitation skill scores. *Wea. Forecasting*, **34**, 3–13, https://doi.org/10.1175/WAF-D-18-0109.1.

Reiter, P., O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, 2018: Does applying quantile mapping to subsamples improve the bias correction of daily precipitation? *Int. J. Climatol.*, **38**, 1623–1633, https://doi.org/10.1002/joc.5283.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Zhu, Y., and Y. Luo, 2015: Precipitation calibration based on the frequency-matching method. *Wea. Forecasting*, **30**, 1109–1124, https://doi.org/10.1175/WAF-D-13-00049.1.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/19/23 08:29 PM UTC