HAL
open science

Precipitation and temperature ensemble forecasts from single-value forecasts

J. Schaake, J. Demargne, R. Hartman, M. Mullusky, E. Welles, L. Wu, H. Herr, X. Fan, D. J. Seo

To cite this version:

J. Schaake, J. Demargne, R. Hartman, M. Mullusky, E. Welles, et al.. Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrology and Earth System Sciences Discussions, 2007, 4 (2), pp.655-717. hal-00298823

HAL Id: hal-00298823

https://hal.science/hal-00298823v1

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hydrol. Earth Syst. Sci. Discuss., 4, 655–717, 2007
www.hydrol-earth-syst-sci-discuss.net/4/655/2007/
© Author(s) 2007. This work is licensed
under a Creative Commons License.

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

Precipitation and temperature ensemble forecasts from single-value forecasts

J. Schaake[1], J. Demargne[1], R. Hartman[2], M. Mullusky[1], E. Welles[1], L. Wu[1], H. Herr[1], X. Fan[1], and D. J. Seo[1]

[1]Office of Hydrologic Development, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, MD, USA
[2]California Nevada River Forecast Center, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, MD, USA

Received: 15 June 2006 – Accepted: 18 December 2006 – Published: 2 April 2007

Correspondence to: J. Schaake (john.schaake@noaa.gov)

Abstract

A procedure is presented to construct ensemble forecasts from single-value forecasts of precipitation and temperature. This involves dividing the spatial forecast domain and total forecast period into a number of parts that are treated as separate forecast events. The spatial domain is divided into hydrologic sub-basins. The total forecast period is divided into time periods, one for each model time step. For each event archived values of forecasts and corresponding observations are used to model the joint distribution of forecasts and observations. The conditional distribution of observations for a given single-value forecast is used to represent the corresponding probability distribution of events that may occur for that forecast. This conditional forecast distribution subsequently is used to create ensemble members that vary in space and time using the "Schaake Shuffle" (Clark et al, 2004). The resulting ensemble members have the same space-time patterns as historical observations so that space-time joint relationships between events that have a significant effect on hydrological response tend to be preserved.

Forecast uncertainty is space and time-scale dependent. For a given lead time to the beginning of the valid period of an event, forecast uncertainty depends on the length of the forecast valid time period and the spatial area to which the forecast applies. Although the "Schaake Shuffle" procedure, when applied to construct ensemble members from a time-series of single value forecasts, may preserve some of this scale dependency, it may not be sufficient without additional constraint. To account more fully for the time-dependent structure of forecast uncertainty, events for additional "aggregate" forecast periods are defined as accumulations of different "base" forecast periods.

The generated ensemble members can be ingested by an Ensemble Streamflow Prediction system to produce ensemble forecasts of streamflow and other hydrological variables that reflect the meteorological uncertainty. The methodology is illustrated by an application to generate temperature and precipitation ensemble forecasts for

the American River in California. Parameter estimation and dependent validation results are presented based on operational single-value forecasts archives of short-range River Forecast Center (RFC) forecasts and medium-range ensemble mean forecasts from the National Weather Service (NWS) Global Forecast System (GFS).

1 Introduction

The National Weather Service (NWS) is implementing a new Advanced Hydrologic Prediction Service (AHPS) (http://www.nws.noaa.gov/oh/ahps). This includes hydrological forecast products to account for the uncertainty in the forecasts, extend forecast lead times out to about a year and improve the accuracy of the forecasts. To help meet these AHPS objectives, the NWS is improving its capability to make ensemble streamflow predictions. Although other methods can be used to quantify uncertainty for specific situations (e.g., Kalman filtering for uncertainty in snowmelt driven water supply forecasts in the western U.S., Day, 1990), the general flexibility ensemble methods provide is needed to satisfy the complex mix of operational and scientific requirements associated with AHPS.

Ensemble methods are essentially a Monte Carlo approach to solving a sequence of non-linear multiple integral equations that cannot be solved analytically. While ensemble methods are most commonly used to quantify uncertainty, they have also been demonstrated to improve forecast accuracy (Georgakakos et al., 2004). There are three primary sources of uncertainty in a river forecast system: future meteorological forcing, initial hydrological conditions, and hydrological modeling uncertainty. Hydrological modeling uncertainty encompasses all the sources of uncertainty associated with translating initial conditions and future forcing into future hydrological fluxes and state variables. Because hydrological systems propagate uncertainty through complex, non-linear processes that operate in space and time, joint space-time distributions of precipitation, temperature and initial conditions, not just their marginal distributions at individual space-time locations, control estimates of uncertainty in hydrological state

and flux variables.

The part of the National Weather Service River Forecast System (NWSRFS) that produces ensemble streamflow forecasts is called the Ensemble Streamflow Prediction (ESP) system. ESP has been used by NWS since the late 1970's (Hirsch et al., 1977; Day, 1985; Smith et al., 1992; Schaake and Larson, 1998). For these early applications the past climatologic variability of precipitation and temperature was assumed to be representative of what might happen in the future and uncertainties in the initial conditions and in the hydrological forecast models were ignored. At that time the ESP acronym meant "Extended Streamflow Prediction". Since then significant improvements have been made in short and medium range forecasts. Accordingly the meaning of the acronym ESP has evolved to become "Ensemble Streamflow Prediction".

This paper describes a methodology to construct precipitation and temperature ensemble members that can be used for ESP and that incorporates the skill of existing operational single-value precipitation and temperature forecasts (currently for periods out to two weeks). These procedures are designed to function in an operational hydrological forecast environment using existing operational meteorological forecast information available at NWS River Forecast Centers (RFCs). The procedures are intentionally made to be simple, have minimum data requirements for parameter estimation and potentially be applicable for any hydrological forecast application. Because of its simplicity it is expected that the technique will have some important limitations that may require future modifications and development of alternative approaches. The approach is being tested in pilot projects at four RFCs.

2 Background

The original application of ESP at the NWS was for long-range forecasts. The assumption was made that atmospheric forcing inputs from historical years were representative of those likely to occur in the future (the climate being considered as stationary). The

precipitation and temperature time series for each historical year produced a single simulation of streamflow that would have occurred if the initial conditions in that year were the same as estimated for the current year.

A first step to apply short-term forecast information in the ESP process was accomplished by linearly blending the single-value Quantitative Precipitation Forecast (QPF) and Quantitative Temperature Forecast (QTF) with the climatologic time series of precipitation and temperature (NWSRFS, 2000). This involved application of relative weights on the forecast and the historical data so that the weight assigned to the forecast decreased to zero during the blending period. The forecaster could control the assignment of weights and the duration of the blending period. This simple blending approach does not account for the intermittency of precipitation, nor does it account for variation in the uncertainty of the QPF with forecast value.

Generating probability distributions for precipitation forecasts with short lead times was addressed by Krzysztzyfowicz (1998, and references therein) as part of a Bayesian forecasting system (BFS) that he proposed as a general theoretical framework for probabilistic forecasting for small headwater basins. His BFS decomposes the total uncertainty into input uncertainty and hydrological uncertainty (uncertainty coming from model limitations, parameter values, initial conditions, measurement error, etc.). These two kinds of uncertainties are quantified independently from each other using an Input Uncertainty Processor and a Hydrologic Uncertainty Processor. The quantification of the precipitation and hydrological uncertainties is then integrated into the probabilistic forecast using Bayes theorem (Kelly and Krzysztzyfowicz, 2000; Krzysztzyfowicz and Kelly, 2000; Krzysztzyfowicz and Herr, 2001; Krzysztzyfowicz, 2001). Assumptions underlying this approach make it most appropriate for application to small watersheds. Seo et al. (2000) proposed an approach to use the Probabilistic Quantitative Precipitation Forecasts (PQPFs) created by Krzysztzyfowicz's BFS to produce ensemble precipitation space-time series from which time series ensemble input for ESP could be generated. The experimental implementation of this approach was pursued in parallel with an experimental implementation of the BFS at the Ohio RFC. Operational limita-

tions of the PQPF procedures underlying the BFS ultimately led to an ending of both experiments.

Clark and Hay (2004) applied model output statistics (MOS) techniques (Glahn and Lowry, 1972) to downscale ensemble mean forecasts from retrospective forecasts for a 40-year period a fixed version of NCEP's Global Forecast System (GFS) to make probabilistic forecasts of daily precipitation and maximum and minimum temperature at the location of each station in the NWS cooperative network. The GFS forecast model was known at the time as the Medium Range Forecast (MRF) System and the re-forecasts were made for an 8-day forecast period every 5 days as part of the NCEP "Reanalysis" project (Kalnay et al., 1996; Kistler et al., 2001)

The forecast uncertainty associated with single-value forecasts is scale dependent. This is especially true for precipitation forecasts. The forecast uncertainty in the single-value precipitation forecast depends on the space and time scale of the forecast, and the verification statistics are also scale-dependent (Tustison et al., 2001; Weygandt et al., 2004). For example, 24-h precipitation forecasts for a given location are more skillful than 6-h forecasts for parts of the same 24-h period. This is because it matters less to determine when precipitation occurs for a 24-h forecast than for a 6-hour forecast. Similarly the forecast for average precipitation over a large area is more skillful that forecast for parts of the area.

With the advent of the AHPS it is essential for NWS to find a practical way to use short and medium term single-value QPF and QTF information to produce ensemble input forcing for ESP. Although ensemble meteorological forecast systems, both regional and global, are now operational, much remains to be done to effectively remove model biases and to compensate for under-prediction of ensemble spread before the output from ensemble meteorological forecast systems can be used for input to ESP. Furthermore, an appropriate role for the human forecaster needs to be developed to preserve the value added by the forecaster to short-term QPF. Therefore, the following procedures are being developed to generate ensemble precipitation and temperature forecasts. These procedures can use existing operational single-value QPF and QTF.

They also can use the ensemble mean forecasts from ensemble precipitation and temperature forecast systems.

3 Methodology

The National Weather Service River Forecast System (NWSRFS) prepares future precipitation and temperature ensemble members for input to ESP using an Ensemble Pre-Processor (EPP). The methodology presented here is used in a new EPP component that transforms time series of single-value QPF and QTF into corresponding ensemble forecasts of precipitation and temperature. The methodology was developed to be as simple as possible, involve as few parameters as possible and use existing single-value forecasts. Although ensemble forecasts are now produced operationally, the uncertainty information in them needs further investigation for forecast members to be made reliable enough for operational ESP application at this time.

The methodology involves two steps. First, the single-value QPF and QTF time series are processed to produce a corresponding set of conditional distributions of precipitation and temperature values that might occur. Then, these conditional distributions are used to assign values to ensemble members using a procedure known as the "Schaake Shuffle" (Clark et al., 2004). This procedure reassigns to historical precipitation and temperature values new values derived from the forecast conditional distributions. It assures that the probability distributions corresponding to the ensemble members are the same as the conditional distributions corresponding to the QPF and QTF. It also assures that the space-time rank correlation structure in the historical data is preserved in the generated ensemble members.

Forecast uncertainty is space and time-scale dependent. For a given lead time to the beginning of the valid period of a forecast, it depends on the length of the forecast time period and the spatial area to which the forecast applies. Although the "Schaake Shuffle" procedure applied to the forecast conditional distributions for each time step may preserve some of this scale dependency, it may not be sufficient to preserve some

of the multi-scale temporal forecast uncertainty without additional constraint. Accordingly, additional aggregate forecast periods are defined as accumulations of different base forecast periods that correspond to the individual time steps. Probability forecasts are also made for these periods and the distribution of aggregate values of the ensemble members is constrained by the corresponding probability forecast. For example, if the forecast time step were 6-h, additional aggregate periods could be constructed for the four 6-h periods for each of the first several days. Further aggregation could be done for periods of two days, three days, etc. Together, the base periods and the additional aggregate periods comprise a set of events. By constructing additional events for aggregate forecast periods, the skill of forecasts over multiple periods of time can be preserved even though there may be very little skill in predicting exactly what may occur during base events with long lead times.

Ensemble members (for a given location) form a matrix with forecast time step on one axis and member number on the other. This matrix is initialized with historical observations corresponding to the current forecast period. The time axis is augmented with additional aggregate forecast periods. But defining additional events to constrain the final ensemble creates an over-determined situation. The ensemble member values generated by the Schaake Shuffle for the base events may not be consistent with the corresponding ensemble members for the aggregate periods. To resolve this conflict, the total set of base and aggregate events is processed in a sequence governed by increasing forecast skill as measured by the forecast-observation correlation parameter for each event. The "Schaake Shuffle" is applied at each step of the sequence using the current data in the ensemble matrix to create an updated ensemble matrix. At the end of the sequence when all events have been processed, the ensemble matrix contains the final set of ensemble members that preserves the temporal scale dependency in forecast uncertainty and skill. The sequential procedure assures that the highest skill events have the greatest influence on the final ensemble values. For now, the spatial scale dependency of forecast uncertainty and skill is controlled by the Schaake Shuffle. It is not clear if additional spatial aggregation of forecast events is required. This will

be examined in future studies.

3.1 Marginal distribution of forecast events

The first step in the ensemble generation procedure is to construct conditional distributions for each forecast period. Slightly different approaches are used to estimate probability distributions for precipitation and temperature because precipitation is intermittent and highly skewed whereas temperature distributions are nearly Gaussian and do not have an intermittent component. The approach to estimate temperature distributions is simpler so it will be presented first.

3.1.1 Temperature distributions

NWS hydrological forecast models are programmed to operate at time steps of multiples of 1-h and factors of 24 h. Most applications have used 6-h time steps but future applications, with steps as short as 1-h, are expected. Although temperature time series at the model time step are required to operate the model, most long-term historical observations of temperature are not made at model time steps. Rather, they are observations of daily maximum (TX) and minimum temperature (TN). Similarly, most historical short term Model Output Statistics (MOS) temperature forecasts have been made for maximum (FTX) and minimum (FTN) temperature. Accordingly, historical analyses of temperature at model time steps for NWS hydrological model applications are made using daily maximum and minimum temperature values. To generate temperature ensembles at the model time step, daily TX and TN ensembles are first generated. Then, ensemble temperature members at model time steps are computed from the corresponding TX and TN ensemble members using the same interpolation procedures used to estimate temperature time series from historical TX and TN observations. Accordingly, the base time step for daily maximum and minimum temperature ensemble forecasts is one day.

To generate TX and TN ensemble members the total forecast period is first divided

into a number of daily base periods. Then, to account for temporal scale dependency of uncertainty, a number of additional aggregate periods are created that represent averages of TX and TN over different combinations of base periods depending on the specific application. The set of base and aggregate periods define a set of events. These events involve a joint probability relationship between values of FTX or FTN, X, and corresponding TX or TN observations, Y.

Climatological distributions of maximum and minimum temperature forecasts (FTX and FTN) and observations (TX and TN) tend to be nearly Normal. Although a transformation could be applied to the temperature forecasts and/or observations, that does not appear to be necessary at this time. Therefore, the joint distribution of temperature forecasts (FTX or FTN) and observations (TX or TN)) is assumed to be Bivariate Normal. The conditional distribution of values of TX or TN that might occur, given FTX or FTN for each event is, therefore, Normal.

The mathematical representation of the daily maximum and minimum ensemble temperature forecast procedure is as follows. For an event corresponding to a given lead time (in days), a given location and a valid period for a given number of days, let X represent the forecast FTX or FTN during the valid period and let Y represent the corresponding TX or TN. (For aggregate forecast periods the average values of FTX or FTN and TX or TN during the aggregation period are used). Conditional distributions for each even are estimated separately and independently. The joint distribution of X and Y is assumed to be Bivariate Normal. The forecast distribution for Y, given a forecast X is the conditional distribution of Y given X which can be extracted from the joint distribution of X and Y. Let $f_X(x)$ be the probability density function (pdf) of X. The cumulative distribution function (cdf) of X is $F_X(x)$. The mean and variance of $f_X(x)$ are μ_X and σ_X^2. Similarly, let $f_Y(y)$ be the pdf of Y, and $F_Y(y)$ the cdf of Y. The mean and variance of $f_Y(y)$ are μ_Y and σ_Y^2. Let the bivariate pdf of X and Y be $f_{XY}(x, y)$ which is modeled as Bivariate Normal with the five following parameters: means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and the correlation parameter ρ_{XY}. It follows that the desired conditional distribution of Y, given the forecast realization x, $f_{Y|X}(y|x)$ is a Normal

distribution with mean and variance given by:

$$\mu_{Y|X} = \mu_Y + \rho_{XY}(\sigma_Y/\sigma_X)(x - \mu_X) \quad (1)$$

$$\sigma^2_{Y|X} = \sigma^2_Y(1 - \rho^2_{XY}) \quad (2)$$

We have found, as expected, the correlation parameter ρ_{XY} to be dependent on the lead time to the beginning of the valid forecast period and on the duration of the valid period of the forecast. It also depends on geographic location and season of the year. The correlation parameter ρ_{XY} also serves as a measure of the forecast skill, being 1 for a perfect forecast and 0 for a completely unskilled forecast. The marginal pdfs, $f_X(x)$ and $f_Y(y)$, represent the climatologies of X and Y, respectively. Equations (1) and (2), show that as the skill of the forecast decreases (i.e., as ρ_{XY} goes to 0), the conditional density $f_{Y|X}$ approaches the marginal density f_Y. Indeed, with ρ_{XY} equal to 0, the results are independent of the current single-value forecast since the temperature distribution $f_{Y|X}$ is the Normal approximation to climatologic distribution f_Y. With ρ_{XY} equal to 1 the conditional distribution of Y converges to a Dirac delta function located at the bias-adjusted value of the single-value forecast: $y = \mu_Y + (\sigma_Y/\sigma_X)(x - \mu_X)$ (i.e. there is no uncertainty).

For both the maximum and the minimum daily temperature, two marginal distributions and one correlation parameter are needed for each forecast lead-time and spatial location: $F_X(x)$ the marginal distribution of forecasts, $F_Y(y)$ the marginal distribution of observations, and ρ_{XY} the correlation coefficient between forecasts and observations. Parameters of these distributions are estimated from an archive of forecast and observation pairs as explained below.

3.1.2 Precipitation distributions

Ensemble precipitation members are generated from a time series of future precipitation forecasts (QPF). Each ensemble member contains basin average precipitation values (MAP) at the model time step for the total forecast period. Although the time

steps of the QPF and MAP values do not have to be the same, the methodology does condition the distribution of MAP values on appropriate corresponding values of QPF. The approach is to divide the total forecast period into a number of base periods, one for each model time step and a number of additional aggregate periods that represent accumulations of precipitation over different combinations of base periods depending on the specific application. The set of base and aggregate periods define a set of events. These events involve a joint probability relationship between a precipitation forecast, X and corresponding precipitation observation, Y. Values of X and Y depend on values of QPF and MAP. For base events there is a one-to-one correspondence between X and QPF and between Y and MAP if the time step of the QPF is the same as the model time step. But each event (both base and aggregate) has a prescribed relationship between values of QPF and X and values of MAP and Y. Subsequently, these relationships are used to create ensemble members from a set of conditional distributions for $Y|X$, there being one conditional distribution for each event.

The approach to estimate the conditional distribution for precipitation events, $f_{Y|X}(y|x)$, given a single-value QPF, X, is similar to the approach for temperature but considers the intermittent property of precipitation. Accordingly, let $f_X(x)$ be the marginal climatological pdf of X. Let $f_Y(y)$ be the marginal climatological pdf of Y. To account for the probability associated with the forecast precipitation amount being zero, $f_X(x)$ is modeled as a mixture of two random variables. One component corresponds to the likelihood for rain not to occur; the other to the amount of rain if it were to occur. This mixture has a pdf with discrete (represented by a Dirac Delta function) and continuous components. The mixture pdf is:

$$f_X(x) = (1 - p_X) \quad \delta(x) + p_X f_{XC}(x|x > 0) \tag{3}$$

where p_X is the probability of occurrence of an amount of precipitation exceeding a prescribed threshold (e.g. more than 0.254 mm for observed point precipitation) and δ is the Dirac delta function. The continuous part is represented by the conditional pdf for precipitation, given that forecast precipitation occurs (i.e., $x>0$). Hence, from Eq. (3),

the cumulative distribution function F_X has the following form:

$$F_X(x) = 1 - p_X, if\, x = 0, \text{ and}$$
$$F_X(x) = 1 - p_X + p_X F_{XC}(x|x > 0) \text{ otherwise}. \quad (4)$$

$F_X(x)$ is a mixed distribution that assigns a probability mass of (1- p_X) to events with no forecast precipitation and spreads the remaining probability mass of p_X to all other events. The density $f_Y(Y)$ and cumulative distribution function $F_Y(Y)$ have similar forms based on p_Y, the probability that an amount of observed precipitation above the prescribed threshold occurs (i.e., y>0). (Note the thresholds for X and Y do not have to be the same).

In order to model (i.e. parameterize) the joint distribution F_{XY} the climatological marginal distributions F_X and F_Y are used to transform variables X and Y to variables U and V, each having marginal Standard Normal distributions. The joint distribution between U and V is assumed to be Bivariate Standard Normal (BSN). The normal quantile transform (NQT), described by Kelly and Krzysztofowicz (1997), is applied as follows to map the non-Normal variables X and Y into standard Normal variables U and V, respectively, with the inverse Q^{-1} of the standard Normal cdf:

$$u = Q^{-1}(F_X(x)) \text{ if } x > 0, \quad (5)$$

$$v = Q^{-1}(F_Y(y)) \text{ if } y > 0. \quad (6)$$

If $x = 0$, the value of u is unknown but is less than u_0 where $Q(u_0) = 1 - p_X$. Likewise, if $y = 0$, the value of v is unknown but less than v_0 where $Q(v_0) = 1 - p_y$.

From Eqs. (5) and (6), realizations of X and Y corresponding to realizations of U and V can be obtained from the inverse transforms:

$$x = F_X^{-1}(Q(u)) \text{ if } u > u_0; x = 0, \text{ otherwise}. \quad (7)$$

$$y = F_Y^{-1}(Q(v)) \text{ if } v > v_0; y = 0, \text{ otherwise}. \quad (8)$$

Applying the inverse transform to a variable that has been transformed via the NQT will result in the original variable.

The joint pdf of U and V, $b(u, v)$, is assumed to be Bivariate Standard Normal with parameter ρ_{UV}, the correlation coefficient between U and V in standard normal space. The parameter ρ_{UV} is a measure of forecast skill, being 1 for a perfect forecast and 0 for a completely unskilled forecast.

In the present approach the probability of precipitation corresponding to a given single-value QPF, X, is implied by the joint distribution between U and V and is not estimated explicitly for a given value of X. Because the sample sizes of precipitation forecast and observed pairs are often too small to enable the user to estimate separately the bivariate properties of discrete and the continuous components of the mixtures in the untransformed space of X and Y, we prefer to make this transformation from X and Y to U and V together with the assumption that U and V are Bivariate Standard Normal (BSN) to keep the model as parsimonious as possible until larger sample sizes of X and Y are more generally available and we have more experience with alternative representations of X and Y. Moreover, our experience with dependent validation of this approach has shown that this methodology is very promising. Others (Bell, 1987; Bellerby, 2005) have used similar approaches to transform precipitation into normally distributed random variables.

The conditional distribution $F_{Y|X}(y|x)$ is used to estimate the cumulative distribution of observed precipitation conditioned by a given forecast. Two cases need to be studied when assessing this conditional distribution: (1) the forecast realization x is greater than zero; (2) the forecast realization x equals zero.

If x is greater than zero, the realization u corresponding to x is calculated using Eq. (5). Since U and V are BSN deviates, the conditional distribution $B_{V|U}(v|u>u_0)$ is a Normal distribution with mean and variance given by:

$$\mu_{V|U} = \rho_{UV} u \tag{9}$$

$$\sigma^2_{V|U} = (1 - \rho^2_{UV}) \tag{10}$$

It follows conditional values of y corresponding to $x > 0$ are implied by $F_{Y|X}(y|x>0) = B_{V|U}(v|u>u_0)$ and are given by $y = F_Y^{-1}(Q(B_{V|U}^{-1}(v|u>u_0)))$.

If the forecast realization x equals zero, the exact value of u corresponding to x is unknown. It is only known that u is less than or equal to u_0, where $u_0 = Q^{-1}(1 - p_X)$. The conditional probability distribution $B_{V|U}(v|u\leq u0)$ is:

$$B_{V|U}(v|u \leq u_0) = \int_{-\infty}^{v}\int_{-\infty}^{u_0} b(u,v)\,du\,dv \Big/ \int_{-\infty}^{\infty}\int_{-\infty}^{u_0} b(u,v)\,du\,dv \tag{11}$$

Equation (10) can be expanded as follows:

$$B_{V|U}(v|u \leq u_0) = \int_{-\infty}^{v}\left(\int_{-\infty}^{u_0} b(u|v)\,du\right) b(v)\,dv \Big/ \int_{-\infty}^{\infty}\left(\int_{-\infty}^{u_0} b(u|v)\,du\right) b(v)\,dv \tag{12}$$

where $\int_{-\infty}^{u_0} b(u|v)\,du = B(u_0|v)$ is normal with mean $\mu_{U|V} = \rho_{UV}v$ and variance $\sigma_{U|V}^2 = (1 - \rho_{UV}^2)$.

It follows that conditional values of y corresponding to $x=0$ are implied by $F_{Y|X}(y|x=0) = B_{V|U}(v|u\leq u_0)$ and are given by $y = F_Y^{-1}(Q(B_{V|U}^{-1}(v|u\leq u_0)))$.

3.2 Ensemble construction

Ensemble values are constructed using the Schaake Shuffle (Clark et al., 2004) by re-scaling historical observed events so that ensemble member values are distributed with the conditional distribution If there are n historical values, a set of n values of Y are derived from $F_{Y|X}(y|x)$ using stratified sampling. For each event, the largest value in the climatological record with $r_k = n$ is re-assigned to the highest ensemble forecast value, y_n; the smallest event with $r_k = 1$ is re-assigned to the lowest ensemble forecast value, y_1. Hence the ranks of the climatological values match the ranks of the ensemble

members. Applying this technique for each event in the forecast period creates ensemble forecasts that preserve the space-time patterns of the climatological values. Clark et al. (2004) applied this method of ensemble member construction for temperature and precipitation and called it the Schaake Shuffle. They examined the space-time variability in local-scale ensemble forecasts of precipitation and temperature in four river basins in the USA. They found that the reconstruction methodology preserves the spatial correlation between all station pairs for all variables in each of the study basins. They also found that the reconstruction approach is capable of preserving the climatological correlation between variables (e.g., correlations between precipitation and temperature). Clark and Hay did not use additional aggregate events as are used here to preserve the conditional correlation structure implied by the multi-scaling nature of forecast uncertainty at different temporal scales.

3.2.1 Temperature

The following re-scaling process is used for both maximum temperature and minimum temperature as follows:

1. First the daily maximum/minimum temperature values for the current day in each of n years in the climatologic record are ranked in the increasing order. The number of ensemble members to be constructed is n. Each year, k, is associated with a rank r_k according to the daily maximum/minimum temperature value that occurred for that year relative to other years.

2. Second, the conditional cdf $F_{Y|X}(y|x)$ is partitioned into n equal intervals of probability equal to 1/n.

3. Then for each interval i varying from 1 to n, the expected value of Y given that Y is in that interval, noted y_i, is computed.

4. Finally, each of the n values of y_i are assigned to replace the historical value for year k by associating i and k so that $r_k = i$.

The distribution mapping process, which is used for both maximum temperature and minimum temperature, generates an ensemble of maximums and an ensemble of minimums for each day in the forecast period. The final temperature ensemble is computed from those two sets using a diurnal cycle to temporally interpolate between the maximum and minimum ensemble members to produce time series of temperature values at the model time step (e.g., 6-h time step).

3.2.2 Precipitation

A similar process is used to construct precipitation ensemble members for each time step of the forecast period although there are differences to account for the intermittency of rain events. The distribution mapping process for precipitation works as follows:

1. First, precipitation values in each of n years in the climatologic record are ranked in the increasing order for each precipitation event. The number of ensemble members to be constructed is n. Each year, k, is associated with a rank r_k according to the precipitation value that occurred at the current time for that year relative to other years. In years when no rain occurred, a rank is currently assigned randomly with none to exceed the smallest non-zero event rank.

2. Second, the conditional cdf $B_{V|U}(v|u)$ is partitioned into n equal intervals of probability equal to 1/n.

3. Then for each interval i varying from 1 to n, the y_i expected value of Y (corresponding to the value F_Y^{-1} $(B_{V|U}(v|u))$ given that V and Y are in the interval i is computed.

4. Finally, each of the n values of y_i is assigned to replace the historical value for year k by associating i and k so that $r_k = i$.

This distribution mapping process is applied for each time event. The sequence of events is ordered in increasing order of the magnitude of the ρ_{UV} correlation parameters. Values of Y for aggregate events are used to constrain the corresponding aggregate values of Y of the corresponding values of Y for individual model time steps. This preserves the appropriate space and time relationships between different precipitation and temperature events. This is crucial for the ESP system that ingests precipitation and temperature ensembles to generate streamflow ensembles.

4 Example application to North Fork American River, California

An example application of the short-term temperature and precipitation ensemble generation procedures is illustrated using data from a high elevation sub-area of the American river basin in the Sierra Nevada mountains of northern California. The American river basin was chosen for several reasons. The basin has complex mountainous terrain so we can demonstrate that the technique is applicable to mountainous areas. Precipitation is highly seasonal, occurring mainly during the cool season between November and April. Precipitation and temperature climatologic statistics vary during the cool season in this basin so we can test how well the parameter estimation procedures can represent this variability. The best historical archive of short range QPF forecasts exists for locations in the CNRFC (California-Nevada RFC) area, including 6-hour QPF amounts for all 6-h periods for lead days 1 to 5.

The American River flows into the Sacramento River at Sacramento, CA. Approximately 40 km upstream of Sacramento is Folsom Reservoir, a multi-purpose water management project operated by the Bureau of Reclamation. Three forks of the American River (North, Middle, and South Forks) drain approximately 4800 km^2 of mountainous terrain (with elevation values up to 3000 m) and join to provide the inflow to Folsom Lake. This basin is characterized by orographic precipitation patterns associated with steep terrain barriers and with snow in the high elevations (typically above 1500 m) (Carpenter and Georgakakos, 2001).

The results presented here are for the upper elevation zone of the North Fork of the American river basin above Folsom Lake. Precipitation data associated with this sub-area are labeled nfdc1huf; temperature data are labeled mfac1luf. Both temperature and precipitation "observations" used in this example are analyzed values that represent area averages and are based on observations from several point locations.

4.1 Forecast data sources

Two sources of single-value precipitation and temperature forecasts are used in this example application. One is RFC short range (1–5 day) forecasts that are produced at the CNRFC from a range of guidance products including forecasts from the Hydrometeorological Prediction Center (HPC), Weather Forecast Offices (WFOs), Regional and Global models and Model Output Statistics (MOS) products. These are available only for the last few years. Precipitation forecasts are for 6 h periods for lead times of 1 to 5 days for subarea nfdc1huf. Daily maximum and minimum temperature forecasts are available for lead times of 1 to 7 days for subarea mfac1huf . The other source of forecasts is medium range (1–14 day) ensemble mean forecasts from a fixed version of the Global Forecast System (GFS) ensemble forecast model (Hamill, 2004). These are available from January, 1979 to the present. They are available for 12 h time steps. Only the ensemble mean forecasts are used because the uncertainty information in the GFS ensemble forecasts is not reliable (Whitaker and Loughe, 1998).

4.2 Observations data sources

Two sources of precipitation and temperature observations data are used in this example application. One is RFC archives of operational precipitation analyses for nfdc1huf and temperature analyses for mfac1luf. The other is analyses of historical precipitation and temperature data from the National Climatological Data Center (NCDC) for nfdc1huf and mfac1luf. Different observing systems are used to produce the operational and historical analyses.

4.3 Design of the example application

The objective of this application is to illustrate the methodology and to demonstrate the potential for it to be useful for generation of ensemble forcing for hydrological ensemble prediction using single-value atmospheric forecasts. Clearly, this methodology has many limitations. These can be expected to depend on climatology, the accuracy of the single-value forecasts, amount of data available to estimate parameters and on basin hydrological characteristics. It is not possible in a simple example to evaluate these limitations.

Example ensemble temperature forecasts are presented for mfac1luf; example precipitation forecasts for nfdc1huf. In order to illustrate and compare ensemble forecasts using both RFC and GFS single-value forecasts, the period of data used is limited to the last few years when both RFC and GFS forecasts are available. All forecasts in this example are for lead times of 1 to 5 days. All ensemble forecasts use the same aggregate forecast periods to account for the effects of temporal scale dependency in forecast uncertainty.

Given the limited objectives of this example, the limited amount of data available to estimate parameters and the desire to show that the methodology can account for seasonal differences in climatology and single-value forecast characteristics; the validation results presented will be for the same period as used to estimate model parameters. This has the advantage that the ensemble generation procedures should be fully calibrated relative to the climatology of the observations used to do the evaluation. Any biases in the examples presented here occur because of limitations in the way the data are represented by the procedures. This is an important starting point for evaluating the methodology.

5 Temperature ensemble example for mfac1luf

The temperature ensemble example illustrates the application of RFC daily maximum and minimum temperature forecasts and the application of GFS 12 h temperature forecasts to produce probabilistic daily maximum and minimum temperature forecasts for mfac1luf. Daily maximum and minimum temperatures GFS 12 h temperature forecasts are associated with daily maximum and minimum temperature forecasts for mfac1luf. A common set of temperature values based on gage observations are used to estimate parameters and to evaluate both the RFC and GFS based TX and TN probabilistic forecasts. This common observations data set is based on operational TX and TN observations if they exist and on NCDC observations otherwise.

The first part of the example presents some parameter estimation results. The second part presents selected forecast verification results using the same data as were used to estimate the parameters. Although the GFS forecasts extend out to fourteen days of lead time, the examples presented here are for only the first five days of forecast lead time when both RFC and GFS forecasts are available so that comparisons between the results can easily be made. Also, the historical period used is for calendar years 1998 to 2002. This is the common period for which all required forecasts and observations were available. This period is too short to support a split sample approach to parameter estimation and forecast validation. Therefore the validation results are for the same period as used for parameter estimation and do not represent independent validation of the procedures. But they do test whether the proposed probabilistic representation of the forecasts can represent the distribution characteristics of the observations when the procedures are well calibrated and when the climatological characteristics of the validation period are the same as used for calibration.

5.1 Temperature parameters

Parameters for the temperature ensemble procedures are estimated using single-value forecast and observation pairs of daily maximum and minimum temperature values

(FTX/FTN and TX/TN). Ideally historical archives of these data should be for many years with longer archives required as the correlation parameter between forecasts and observations approaches zero. If the correlation parameter were equal to zero, the conditional distribution would be the climatological distribution which does not depend on the forecast. In that case, a long archive of 30 or more years is preferred. If the correlation were nearly equal to one, only a short archive would be needed. The question of required archive length is beyond the scope of this work. It is being investigated and will be presented in future papers.

The period used for the analysis in this example is the 4-year period 1998 to 2001 when RFC and GFS forecasts and observations were available in the data base. For both daily maximum and minimum temperatures, the joint distribution for each temperature event has five parameters (μ_X,σ_X,μ_Y,σ_Y, and ρ_{XY}). The method of moments is used to estimate μ_X,σ_X,μ_Y and σ_Y directly from archived temperature values. The value of ρ_{XY} is estimated as the coefficient of correlation between forecasts and observations. To assure that values of these parameters vary gradually from day to day during the year and to minimize the required length of historical archive, all data in a window centered on a given day are used to estimate parameters. This data pooling process increases the sample size in order to better estimate the distribution parameters and to remove random sampling noise for extreme events. In the examples presented below the total width of the window used is 91 days with 45 days before and after the target day. This was done for time steps of every 5 days throughout the year. Daily parameter values were interpolated from the 5-day values. The length of the time window was chosen subjectively to balance the need for data with the need to respect the seasonality of the parameter values.

Values of average daily minimum temperature observations (μ_Y) and forecasts (μ_X) are illustrated in Fig. 3. There are four plots in this figure. Figure 3a is for values of μ_Y for RFC-based forecasts. Figure 3b is for values of μ_X for RFC-based forecasts. Figure 3c is for values of μ_Y for GFS-based forecasts. Figure 3d is for values of μ_X for GFS-based forecasts. Each plot shows how average minimum temperature forecasts

or observations vary with season throughout the year (vertical axis) for each of eight forecast periods (horizontal axis). Units are degrees Celsius. The first five forecast periods correspond to forecasts for future days one to five. Forecast periods six to eight represent average forecasts for different aggregations of the forecasts for the first five days. Period six is the average of forecasts for days one and two. Period seven is the average of forecasts for days three to five. Period eight is the average of forecasts for all five days.

Figures 3a–d show there are strong seasonal changes in the averages but there is very little change in the averages with forecast period at a given time of year, as is expected. The average TN observations for RFC and GFS forecasts in Figs. 3a and c are the same because a common set of observations data were used for these figures. The average RFC TN forecasts in Fig. 3b are much closer to the corresponding average TN observations (Fig. 3a) than are the average GFS TN forecasts in Fig. 3d to the corresponding observations in 3c. This is because the GFS TN forecasts are for a 2.5 degree grid element that includes mfac1luf. But mfac1luf is in the Sierra Nevada mountains and has a much higher elevation than the average elevation of the corresponding GFS grid element. One of the implications of Fig. 3 is that it is essential to remove biases from atmospheric model forecasts before they are used as input to hydrological models. As will be shown below the biases in both the RFC and GFS forecasts that are indicated in Fig. 3 are removed by the proposed procedures.

Figures 4a–d show the standard deviations (σ_X and σ_Y), in degrees Celsius, corresponding to the averages presented in Figs. 3a–d. The observed standard deviations for RFC and GFS based forecasts in Figs. 4a and c are identical because the same data were used to produce these plots. The RFC forecasts have much smaller standard deviations than observed. The standard deviations of the GFS forecasts tend to be slightly larger than the observed standard deviations. There is evidence of temporal scale dependence in the standard deviations of the temperatures because the observed standard deviations for forecast periods 6 to 8 are less than the corresponding standard deviations of the daily values for periods one to five. The RFC forecasts have

much smaller standard deviations than observed. There is some evidence that forecast standard deviations decrease with lead time in forecast periods 1–5. A decrease in forecast standard deviation is consistent with the fact that the standard deviation of an optimal single-value minimum variance estimator is less than the standard deviation of the predictand as a function of the correlation coefficient between them, as given in Eq. (2). Nevertheless, the procedures proposed here remove the effects on the temperature probability forecasts of any biases or inconsistencies in the single-value temperature forecasts used to drive the procedures.

Figures 5a and c show the correlation coefficient between single-value forecasts and observations. Figure 5a corresponds to RFC single-value forecasts; Fig. 5c, to GFS single-value forecasts. For both forecasts the correlations are higher in spring and fall than in summer and winter. The GFS single-value temperature forecasts are generally more highly correlated with observations than the RFC temperature forecasts at this location. Figures 5b and d present the corresponding correlation coefficients for the ensemble mean of the probabilistic minimum daily temperature forecasts generated by the proposed preprocessing procedures. The ensemble mean correlations with observations are almost identical to the single-value forecast correlations for both RFC and GFS forecasts. This means that the proposed preprocessing procedures preserve the skill inherent in the single-value forecasts. Figures 5a to d show that the correlation coefficient tends to decrease with increasing lead time (forecast periods 1–5) and is highly temporally scale dependent (forecast periods 6–8). In fact the correlation between the mean minimum temperature for days 1–5 (forecast period 8) is almost as great as the correlation for minimum forecasts on day 1.

5.2 Temperature ensemble validation

Owing to the shortness of historical record of short-term forecasts of precipitation and temperature (see below for details), only dependent validation could be carried out. The same data set used for parameter estimation is also used for validation (1988–2002). The validation statistics presented here only measure how well the model pro-

cedures fit the calibration data and aim at establishing the plausibility of the approach. To evaluate the temperature ensemble forecasts generated with this methodology, daily maximum and minimum temperature ensembles were generated and compared with the corresponding observations. For both maximum and minimum temperature, the RFC or GFS single-value forecast for a given day was used together with the corresponding joint distribution parameter values to estimate the conditional probability distribution for that day and then derive ensemble members for maximum and minimum daily temperatures. This produced a forecast time series of maximum and minimum temperatures for each ensemble member. Each time series was interpolated to a corresponding time series of 6 h temperature values that are required for input to the hydrological forecast model. In this section, only validation statistics for the daily minimum temperatures are presented.

A few verification statistics have been selected to be presented in this paper, based on the forecast verification measures described by Wilks (1995), Jolliffe and Stephenson (2003) and Herbach (2000). Also the verification statistics obtained for the generated ensemble means are compared to the ones for the input single-value forecasts in order to evaluate the quality of the probabilistic forecast against the quality of the single-value forecast. Results are presented below for both RFC and GFS single-value forecasts.

Figures 6a to d present bias statistics for minimum daily temperature forecasts. The units of bias are degrees Celsius. Both RFC and GFS single-value forecasts have seasonally varying biases (Figs. 6a and c). Biases at a given time of year are not strongly dependent on forecast period. RFC minimum temperature forecasts have positive biases as large as about 2 degrees Celsius in March and April. Biases in GFS single value forecasts vary between 2 and 4 degrees Celsius higher than observed. One cause of this bias in GFS forecasts is that the average elevation of the GFS grid element is much lower than the average elevation of the mfac1luf target area. The corresponding average minimum temperature from the ensemble mean of the probabilistic forecasts generated from the single-value RFC and GFS forecasts is completely

unbiased as shown in Figs. 6b and d. This is to be expected because the validation period is the same as the calibration period. This is an indication that the preprocessor parameters are well calibrated for this period.

Figures 7a to d present RMS error statistics (in degrees Celsius) of the differences between observations and single-value forecasts (Figs. 7a and c) and ensemble mean forecasts (Figs. 7b and d). The RMS errors of the ensemble mean forecasts are less than the RMS errors of the single value forecasts. The main reason for this is that the systematic biases (in both means and standard deviations) in the single-value forecasts have been removed by the preprocessor procedures. The RMS errors in Figs. 7b and d are substantially smaller than the standard deviations of the observations in Figs. 4a and c. This relationship is consistent with the expectation for RMS errors to be lower than the observed standard deviation, depending on the correlation coefficients presented in Figs. 5a to d. The RMS errors tend to increase with forecast lead time (Periods 1–5). But they also are highly temporally scale dependent, depending on duration of temporal aggregation as well as lead time.

Figures 8a to d present values of the Continuous Rank probability Skill Score (CRPSS) for both single-value and generated ensemble forecasts. The CRPSS is a potentially very useful summary verification statistic. Values of the CRPSS presented in Fig. 8 are derived from values of the Continuous Rank Probability Score (CRPS). The CRPS is equivalent to the integral of the Brier Score (BS) over all possible thresholds of the variable being predicted (e.g. TMIN in this case). Accordingly the units of the CRPS are the same as the predicted variable (i.e. degrees Celsius herein). The CRPS can be applied to both single-value and probabilistic forecasts. In the case of single-value forecasts the CRPS is equal to the mean of the absolute values of the difference between the forecast and the corresponding observation. In the case of probabilistic forecasts the CRPS is a probability-weighted average of all possible absolute differences between forecasts and observations. The CRPS for a perfect forecast is zero. The CRPSS is computed using a reference value of the CRPS to create a scale of values ranging from minus infinity to one. The CRPS reference value used here is

the CRPS corresponding to a single-value forecast equal to the climatological mean minimum daily temperature. The CRPSS for a perfect forecast has a value of one. The CRPSS for a climatological single-value forecast has a value of one. A negative CRPSS suggest a forecast worse than climatology. The CRPS tends to be increased by forecast bias and reduced by the effects of correlation between forecasts and observations. Hersbach (2000) presents two alternative decompositions of the CRPS to account for the relative contributions of bias, resolution (i.e. correlation in this case) and climatology (i.e. no forecast skill) to the CRPS.

Values of the CRPSS in Figs. 8b and c for the ensemble forecasts are much improved over the CRPSS values for the single-value forecasts used as preprocessor inputs in Figs. 8a and c. This happens for two main reasons. First, the CRPSS for a probabilistic forecast with reliable probabilities is better than the CRPSS for a single-value forecast equal to the mean of the probabilistic forecast. Second, the CRPSS for an unbiased forecast is better than the CRPSS for a biased forecast. In Figure 8c the white space is for highly biased single-value GFS forecasts with CRPSS below the lowest contour level of 0.5. CRPSS values for ensemble forecasts based on single-value GFS forecasts (Fig. 8d) are better than CRPSS values for ensemble forecasts based on single-value RFC forecasts (Fig. 8b) because GFS-based forecasts are more highly correlated with observations than RFC based forecasts (see Figs. 5a and c). Values of the CRPSS are shown to be highly temporally scale dependent. Larger CRPSS values obtain for aggregate forecast periods (Periods 6–8).

A summary measure of the reliability of the probabilities associated with the generated probability forecasts is presented in Figs. 9a and c. These figures present the RMS probability error of the daily minimum temperature probability forecasts. They show that the RMS probability error is less than about 0.05 for all forecasts at any time of the year. The RMS probability error is a measure of the information that can be displayed in a so-called "reliability diagram" that compares forecast probability values with the relative frequency of occurrence of the forecast events. In this study, the forecast non-exceedance probability associated the observation for each forecast event was

computed using the corresponding event probability forecast. Non-exceedance probabilities were binned into quartiles and the fraction of events occurring for each quartile was compared to the average non-exceedance probability for that quartile. The RMS difference between these values is the RMS probability error shown in Fig. 9.

6 Precipitation ensemble example for nfdc1hufuf

The precipitation ensemble example illustrates the application of RFC 6-h single-value precipitation forecasts and the application of GFS 24 h ensemble mean precipitation forecasts to produce probabilistic precipitation forecasts for nfdc1huf. Precipitation in this area has a very strong seasonal variation. Most precipitation occurs in the cool season from October through March. There is very little precipitation in the middle of summer. The first part of the example presents some parameter estimation results. The second part presents selected forecast verification results using the same data as were used to estimate the parameters.

6.1 Precipitation parameters

Precipitation parameters also are estimated using forecast and observation pairs. The period used for precipitation analysis in this study is the 5-year period 1998–2002 when both RFC and GFS forecasts and observations were available in the data base. Archived RFC forecasts and 6 h precipitation observations and corresponding MAP analyses begin at CNRFC in October, 1997. The archive of GFS precipitation forecasts begins in 1979. In this study only the ensemble mean 24 h GFS precipitation forecasts are used. The GFS 24 h total forecast is assumed to occur equally in four 6 h periods. The archive of historical MAP analyses (based on NCDC data) begins in 1948. Unfortunately, the statistics of the MAP operational and historical observations analyses for nfdc1huf have important differences. This happens in part because different precipitation gages are used in the operational network and partly because of

differences in the temporal reporting characteristics of the operational vs climatological gages. Because of these differences the MAP precipitation analyses are not combined into a composite data base as was done for temperature data, above. Accordingly, the historical MAP analyses are used to estimate parameters for the GFS forecasts (and validate them) and the archived MAP analyses are used to estimate parameters for the RFC forecasts (and validate them). The common period when all required RFC and GFS precipitation forecasts and observations are available is the 5-year period 1998–2002.

The joint distribution of precipitation forecasts and corresponding observations has marginal distributions that are characterized by separate climatologic marginal distributions of forecasts and observations. The statistics of these marginal distributions of both forecasts and observations must be estimated from historical archives of forecasts and observations. Then, the marginal distributions are used to transform the forecast and observed values to new variables that are assumed have a Bivariate Standard Normal distribution.

The joint distribution of precipitation forecasts and observations is a mixture can be partitioned into four sub-regions as illustrated in Fig. 1. One of these sub-regions contains events when both forecast and observed values are zero; the fraction of the total events in this sub-region is P_{00}. Another sub-region contains events when both forecast and observed values are non-zero; the fraction of the total events in this sub-region is P_{11}. The other two sub-regions contain events with either forecast values of zero and non-zero observed values (P_{01} fraction) or non-zero forecast values with observed values of zero (P_{10} fraction).

The climatologic marginal distributions also are mixtures. Each has three parameters (p_X, $\mu_{X|X>0}$ and $CV_{X|X>0}$) and (p_Y, $\mu_{Y|Y>0}$ and $CV_{Y|Y>0}$) that must be estimated from forecast and observations pairs. Parameters p_X and p_Y account for the intermittent behavior of the observations and forecasts. Parameters $\mu_{X|X>0}$, $\mu_{Y|Y>0}$, $CV_{X|X>0,}$ and $CV_{Y|Y>0}$ allow the method of moments to be used to estimate the parameters of the distributions for the non-zero events. This limited set of parameters was chosen so that

the resulting probabilistic model would be as parsimonious as possible, an important condition considering the limited forecast data available to estimate model parameters. Parameters $CV_{X|X>0}$ and $CV_{Y|Y>0}$are the coefficients of variation of the non-zero values of X and Y and are defined as:

$$CV_{X|X>0} = \sigma_{X|X>0}/\mu_{X|X>0} \text{ and}$$
$$CV_{Y|Y>0} = \sigma_{Y|Y>0}/\mu_{Y|Y>0} \tag{13}$$

The p_X and p_Y parameters of the climatologic marginal distributions are related to the fraction of events in the four sub-regions as follows: $p_X = P_{10} + P_{11}$ and $p_Y = P_{01} + P_{11}$.

Values of X and Y are transformed into BSN variables U and V, respectively, using the corresponding climatologic distributions. Values of U and V also fall into four sub-regions of the joint distribution of U and V as illustrated in Fig. 2. The (u, v) events fall into the same sub-regions in Fig. 2 as the corresponding (x, y) events in Fig. 1. The joint distribution of U and V is characterized by the correlation parameter ρ_{UV}. Estimation of parameter ρ_{UV} is difficult because not all of the values of U and V can be observed: the values $u<u_0$ and $v<v_0$ cannot be observed. Alternative approaches can be considered; ultimately it may be best to tune parameter ρ_{UV} to maximize certain forecast verification statistics. In this study, parameter ρ_{UV} is estimated as the average of simple Pearson product-moment correlation coefficient of the untransformed variables X and Y (including zeros) and the value of ρ_{UV} that best fits the values of P_{10}, P_{11}, P_{01} and P_{11}.

Values of daily precipitation parameters vary gradually from day to day during the year. As with temperature parameters it is not possible to estimate daily parameter values directly from data for only a given day because the archive of forecast and observed pairs is not long enough. A data pooling process is used to better estimate the distribution parameters based on neighboring events. Therefore, for a given day, the seven parameters of the daily precipitation joint distribution are estimated from all the forecast and observed data that have occurred during a time window centered on

that day. The width of this window is variable to assure that the number of non-zero forecasts or observations existing within the window is not smaller than a minimum threshold. Currently, the parameters are estimated every five days using a moving window; daily parameter values are estimated by interpolating linearly between these values to get smoothed parameter values gradually varying during the year.

Regarding the choice of the marginal distribution function, the values of the coefficient of variation determine the distribution function used to fit observed or forecast data. If the coefficients of variation were equal to 1.0, the exponential distribution might be expected to fit the data well since it has a fixed coefficient of variation equal to 1.0. The coefficient of variation of the observations is greater than 1.0 suggesting that a distribution such as the Weibull might be a better choice than the exponential. Another alternative might be a Gamma distribution. All three of these distributions are exactly the same if the coefficient of variation is 1.0. Experience with the demonstration basins has shown that the Weibull distribution is an adequate choice to represent the climatology of the observations. Regarding forecasts, in general it might be expected that the coefficient of variation for forecast values would be less than for observed values. Moreover, it would be expected to decrease toward zero as the skill in the forecast decreases. Accordingly, we have chosen to represent the climatologic distribution of forecast values using a Gamma distribution. The choice of distribution function to fit the sample distributions of forecasts and observations involves a trade-off between the goodness of fit and the computational resources required to use the distribution function. The operational forecast system was developed to support alternative distribution choices. Ultimately, the final choice of distribution should be made based upon a strategy that optimizes forecast verification statistics.

Figure 10 presents average observation and average forecast precipitation statistics. There are four plots in this figure. Figure 10a is for values of the unconditional average for RFC-based forecasts. Figure 10b is for values of the unconditional average for RFC-based forecasts. Figure 10c is for values of the unconditional average for GFS-based forecasts. Figure 10d is for values of the unconditional average for GFS-based

forecasts. Each plot shows how average minimum temperature forecasts or observations vary with season throughout the year (vertical axis) for each of eight forecast periods (horizontal axis). Units are millimeters. The forecasts are for 6hr time steps for forecast lead days 1–5. Forecast periods 1 to 20 correspond to the 20 6 h time steps in the 5 day forecast window. Forecast periods 21 to 28 correspond to average forecasts for aggregate combinations of periods 1–20. Forecast periods 21–25 represent the daily average 6 h precipitation for each of the 5 lead days in periods 1–20. Period 26 is the average for days 1 and 2. Period 27 is the average for days 3–5. Period 28 is the average for days 1–5.

Figures 10a–d show there are strong seasonal changes in the averages but there is very little change in the averages with forecast period at a given time of year, as is expected. The average precipitation observations for RFC and GFS forecasts in Figs. 10a and c are similar but not the same because different sets of observations data were used for these figures. The average RFC precipitation forecasts in Fig. 10b are much closer to the corresponding average observations (Fig. 10a) than are the average GFS precipitation forecasts in Fig. 10d to the corresponding observations in Fig. 10c. This is because the GFS TN forecasts are for a 2.5 degree grid element that includes nfdc1huf. But nfdc1huf is in the Sierra Nevada and subject to much stronger orographic forcing than the corresponding GFS grid element. Figure 10 reinforces the idea that it is essential to remove biases from atmospheric model forecasts before they are used as input to hydrological models. As will be shown below the biases in both the RFC and GFS forecasts that are indicated in Fig. 10 are removed by the proposed procedures.

Figures 11–14 present statistics p_X, p_Y, $\mu_{X|X>0}$, $\mu_{Y|Y>0}$, $\mathrm{CV}_{X|X>0}$, $\mathrm{CV}_{Y|Y>0}$ characterizing the marginal distributions of precipitation forecasts and observations. The intermittent nature of precipitation is characterized, mathematically, in terms of events that are either zero or non-zero. But it is not possible to make accurate measurements of very small short duration precipitation amounts. Therefore a precipitation threshold is introduced to distinguish between wet and dry events. Unfortunately the all of the

statistics , p_X, p_Y, $\mu_{X|X}>0$, $\mu_{Y|Y}>0$, $CV_{X|X}>0$, $CV_{Y|Y}>0$ are very sensitive to the selection of this threshold. This happens because a large fraction of the average precipitation amount is produced by a very small fraction of the total number of precipitation events. As a result the average precipitation amount during a wet event and the corresponding probability of precipitation (1–p) depend very strongly on the precipitation threshold used to define wet and dry events. The conditional coefficient of correlation is also very sensitive to the threshold, increasing sharply as the threshold assumes very small values. It is beyond the scope of this study to optimize how this threshold should be specified. In this study a precipitation threshold of 0.3 mm/6 h is used as a threshold for 6 h precipitation amounts. Lower values were used for average precipitation amounts for daily averages and for averages over several days.

Figure 11 shows the Probability of Precipitation (POP) (1-p_X, 1-p_Y): (a) POP corresponding to single-value RFC forecasts, (b) POP for single-value RFC forecasts, (c) POP for observations corresponding to single-value GFS forecasts, (d) POP for single-value GFS forecasts. There are major differences between values of POP for the MAP observations analyses used for RFC and GFS forecasts. This is partly because of differences in the average precipitation amounts in the two MAP analyses as shown in Figs. 10a and c. But the main difference is in the distribution properties of small precipitation amounts in the two data sets. As the precipitation threshold is increased above 0.3 mm/6 h, the POP statistics in Figs. 11a and c become more similar. POP statistics of the GFS single-value precipitation forecasts in Figure 11d increase with forecast lead time from forecast periods 1–20 with a step increase after every fourth period. This is to be expected because the variance of the precipitation forecasts should decrease as the forecast skill diminishes with increasing lead time. Accordingly, the fraction of wet events above the low precipitation threshold of 0.3mm would be expected to increase. Corresponding RFC forecast POP statistics in Fig. 11b do not have the same pattern. This difference occurs because different procedures are used to produce the RFC forecasts for days 1–3 and for days 4 and 5. Both the RFC and GFS POP statistics are highly temporally scale dependent. POP values for the aggregate forecast periods are

much greater than the pop statistics for the basic 6 h time steps, as is expected.

Figures 12a to d present the conditional average precipitation statistics ($\mu_{X|X}>0$, $\mu_{Y|Y}>0$) for RFC and GFS observations and forecasts. These statistics have very strong seasonal variations. They also have important variations with forecast period. Figures 12a and c show that the conditional average observations are very different between the MAP observations analyses data sets used for RFC and GFS forecasts. The explanation for these differences is the same as given above for the differences between Figs. 11a and c. The conditional average of the RFC forecasts in Fig. 12b is very different from the observations in Fig. 12a. The same is true for the GFS forecasts in Fig. 12d. Also the conditional average of the GFS forecasts is much lower than the conditional average of the RFC forecasts because the GFS grid element does not have as strong orographic forcing as actually occurs in the nfdc1huf area.

Figures 13a to d present the conditional coefficient of variation statistics ($CV_{X|X>0,}CV_{Y|Y>0}$) for RFC and GFS observations and forecasts. These statistics have strong seasonal variations. Figures 13a and c show that the conditional coefficients of variation are very different between the MAP observations analyses data sets used for RFC and GFS forecasts. The explanation for these differences is the same as given above for the differences between Figs. 11a and c. The conditional average of the RFC forecasts in Fig. 13b is very different from the observations in Fig. 13a. The same is true for the GFS forecasts in Fig. 13d. The conditional coefficient of variation is also temporally scale dependent – values for forecast periods 1–20 are different than those for the aggregate forecast periods 21–28.

Figures 14a and c show the correlation coefficient between single-value forecasts and observations. Figure 14a corresponds to RFC single-value forecasts; Fig. 14c, to GFS single-value forecasts. For both forecasts the correlations are higher in the coo; season than in the warm season. The RFC single-value precipitation forecasts are generally more highly correlated with observations than the GFS forecasts for forecast days 1 and 2. Figures 14b and d present the corresponding correlation coefficients for the ensemble mean of the probabilistic precipitation forecasts generated by the

proposed preprocessing procedures. The ensemble mean correlations with observations are almost identical to the single-value forecast correlations for both RFC and GFS forecasts. This means that the proposed preprocessing procedures preserve the skill inherent in the single-value forecasts. Figures 14a to d show that the correlation coefficient tends to decrease with increasing lead time and is highly temporally scale dependent. Correlation coefficients for aggregate forecast periods tend to be much larger than for any of the corresponding 6hr periods.

6.2 Precipitation ensemble validation

Figures 15a to d present bias statistics for precipitation forecasts. The units of bias are degrees millimeters. Both RFC and GFS single-value forecasts have seasonally varying biases (Figs 15a and c) but the biases in the GFS forecasts are much greater than in the RFC forecasts. Biases at a given time of year are not strongly dependent on forecast period. One cause of the bias in GFS forecasts is that the average elevation of the GFS grid element is much lower than the average elevation of the nfdc1huf target area. The corresponding average precipitation from the ensemble mean of the probabilistic forecasts generated from the single-value RFC and GFS forecasts is completely unbiased as shown in Fig. 15b and d. This is to be expected, partly, because the validation period is the same as the calibration period. This is an indication that the preprocessor parameters are well calibrated for this period.

Figures 16a to d present RMS error statistics (in mm) of the differences between observations and single-value forecasts (Figs. 10a and c) and ensemble mean forecasts (Figs. 10b and d). The RMS errors of the ensemble mean forecasts are less than the RMS errors of the single value forecasts. The main reason for this is that the systematic biases (in both means and standard deviations) in the single-value forecasts have been removed by the preprocessor procedures. The RMS errors tend to increase with forecast lead time (Periods 1–20) and they have a diurnal pattern. But they also are highly temporally scale dependent, depending on duration of temporal aggregation as well as lead time.

Figures 17a to d present values of the Continuous Rank probability Skill Score (CRPSS) for both single-value and generated ensemble forecasts. Values of the CRPSS presented in Fig. 17 are derived from values of the Continuous Rank Probability Score (CRPS) in the same way as explained above for temperature forecasts. Values of the CRPSS in Figs. 17b and d for the ensemble forecasts are much improved over the CRPSS values for the single-value forecasts used as preprocessor inputs in Figs. 17a and c. This happens for two main reasons. First, the CRPSS for a probabilistic forecast with reliable probabilities is better than the CRPSS for a single-value forecast equal to the mean of the probabilistic forecast. Second, the CRPSS for an unbiased forecast is better than the CRPSS for a biased forecast. CRPSS values for ensemble forecasts based on single-value RFC forecasts (Fig. 17b) are slightly better than CRPSS values for ensemble forecasts based on single-value GFS forecasts (Fig. 17d) for the first two forecast days. Beyond that the CRPSS values of RFC and GFS forecasts are about the same. Values of the CRPSS are shown to be highly temporally scale dependent. Larger CRPSS values obtain for aggregate forecast periods (Periods 21–28).

Figures 18a to d present values of the Continuous Rank probability Skill Score (CRPSS) for "wet" events only, for both single-value and generated ensemble forecasts. Figure 17 includes all events. Only events where observed precipitation occurred in excess of the precipitation threshold are included in Fig. 18. Values of the CRPSS presented in Fig. 18 are generally lower than in Fig. 17 indicating that part of the skill score in Fig. 17 represents the ability of the forecasts to distinguish between wet and dry events. Values of the CRPSS in Figures 18a and 18c for the single value forecasts show that the raw single-value forecasts do not predict the magnitude of the wet events very well. On the other hand Figs. 18b and d for the ensemble forecasts show substantial improvement in the CRPSS over the values in Figs. 18a and c. Values of the CRPSS for wet events are also shown to be highly temporally scale dependent.

Figures 19a and b present values of the Brier Skill Score (BSS) for the ensemble forecasts of the probability of precipitation. Figure 19a is for ensemble forecasts based

on RFC single-value forecasts. Figure 19b is for ensemble forecasts based on GFS single-value forecasts. BSS values for GFS based forecasts are a little better than for RFC forecasts for the 6 h forecast periods 1–20. BSS values for aggregate forecast periods 21–28 are much better for GFS than RFC based ensemble forecasts. BSS values for GFS based ensemble forecast are highly scale dependent. BSS values for both RFC and GFS based ensemble forecasts are much better in the cool season than in the warm season.

Figures 19c and d present values of the Brier Skill Score (BSS) for the ensemble forecasts of the probability of precipitation occurring in the upper tercile of the distribution of wet precipitation events. Figure 19c is for ensemble forecasts based on RFC single-value forecasts. Figure 19d is for ensemble forecasts based on GFS single-value forecasts. Only observed wet events were used to compute the BSS in Figs. 19c and d and is based on the ensemble probability forecast conditioned on a wet event occurring. These BSS values are higher in the cool season than in the warm season. They are also strongly scale dependent being much larger for aggregate forecast periods than for the 6 h forecast periods. The RFC based ensemble forecasts had higher BSS values than the GFS based forecasts for these large wet events. It is interesting to note that the RFC based ensemble forecasts (Fig. 19c) were much better at predicting the occurrence of large wet events for aggregate forecast periods than at predicting the occurrence of wet vs dry events (Fig. 19a). This may partly be an artifact of how the precipitation threshold was established.

A summary measure of the reliability of the probabilities associated with the generated precipitation probability forecasts is presented in Figs. 20a–d. Figures 20a and b present RMS errors in probability of precipitation forecasts for RFC and GFS ensemble based forecasts. The RMS probability errors in Figs. 20a and b represent the RMS deviation of the reliability diagram from the diagonal. Values of POP were binned into quartiles and the fraction of wet events occurring for each quartile were compared to the average POP for that quartile to compute these RMS errors. The RMS error in the POP forecasts is generally less than about 10 percent except for RFC based

forecasts for 6 h precipitation events. During the summer period there are almost no precipitation events so these results may be sensitive to the small sample size of wet events available for this study. Figures 20c and d present RMS errors in the conditional probability of precipitation amount, given that precipitation occurs, for RFC and GFS ensemble based forecasts. The RMS probability errors in Figs. 20c and d represent the RMS deviation of the cumulative rank histogram from the diagonal. For each wet event the ensemble forecast non-exceedance probability associated with the precipitation observation was computed. If the probability forecasts were reliable the distribution of these non-exceedance probabilities for a large number of events would be uniform. The RMS deviation of the distribution of non-exceedance probabilities for the observed events from a uniform distribution is presented in Figs. 20c and d. Although many of the forecasts have RMS probability errors less than 10 percent, there are some cases where the RMS probability error approaches 20 percent, especially for the GFS based ensemble forecasts in Fig. 20d. The cause of this could be related to the way the precipitation threshold is established because the parameters of the preprocessor are highly sensitive to this and the shape of the conditional distribution of wet events is very sensitive to this.

7 Conclusions and future work

A procedure is presented to construct ensemble forecasts of precipitation and temperature using existing operational single-value forecasts. This procedure involves constructing the joint distribution of forecasts and observations from a historical archive of these data. For a given single-value forecast, the forecast distribution of events that might occur is estimated as the conditional distribution derived from the joint distribution. Ensemble members are subsequently constructed using historical space-time observations of temperature and precipitation.

Historically observed values are replaced with values sampled from the forecast distribution. The replacement procedure matches the ranks of historical and forecast

values. The procedure is operated independently for each spatial location. In the current application these are hydrological sub-areas but the procedure could be applied to grid locations. At each location the procedure is applied independently to a number of events. These events represent both a sequence of individual time steps and a set of optional aggregate events that can be defined to assure that temporal scale dependency properties of forecast uncertainty are preserved in the ensemble members created by the procedure. The marginal distributions of ensemble values for these events are controlled by the current forecast. The procedure to construct precipitation ensembles is similar to the procedure for temperature, except that additional steps are needed to deal with the intermittency of precipitation and with skewness in the distribution of precipitation amounts. This is done using variable transformations to convert precipitation forecasts and observations to new variables that have a bivariate standard normal distribution.

The procedure has been applied to a sub-area in the North Fork of the American River basin, California. Estimated parameters and dependent validation results are presented for minimum temperatures as well as precipitation.

Parameter estimation was based on a data pooling process to get more robust parameters: the parameter values for a given day were computed from events occurring within a time window centered on that day. For temperature, the validation results indicate that daily ensemble forecasts are skillful for the five lead days. Verification statistics for minimum temperature suggest that model assumptions that temperature observations and single value forecasts are distributed according to a bivariate normal distribution are reasonable and that the probability forecasts are both skillful and reliable. Verification statistics for precipitation indicate that there is skill both in the derived probability of precipitation forecast and in the ensemble mean value of the forecast distribution of precipitation amounts given that precipitation occurs. The ensemble procedure aims at reducing the bias of the single-value forecast even if there is still some bias in the ensemble mean. The verification statistics show that there is daily skill in precipitation probabilistic forecasts for at least the first five days. Probability of precip-

itation forecasts and are shown to be reliable but probability forecasts of conditional precipitation amount were not as reliable as the probability of precipitation. It was also shown that probability forecasts for high precipitation amounts (defined as in the upper tercile of observed precipitation values) have skill, even out to five days. It was shown that forecast uncertainty in both minimum temperature and precipitation forecasts is temporally scale dependent. Verification statistics reflect this scale dependency and demonstrate that forecasts for aggregate forecast periods tends to be stronger than for shorter periods. The proposed preprocessor procedures preserve the scale dependency.

Two sources of single-value forecasts were used in this study. One is single-value RFC forecasts that are available for forecast periods up to 5 days for precipitation and up to 7 days for temperature. The other is single-value GFS forecasts for 14 days. For the first few days both RFC and GFS forecasts are available. In this study they were used separately. But it may be possible to produce better ensemble results if either RFC and GFS forecasts were used together in a multivariate approach or if an automatic procedure were used to determine which forecast would be used at any given time of year and for which forecast lead times.

The practical reliability of the procedures proposed here depends on having a long enough historical archive of forecasts and observations to estimate model parameters. Typically, only a few years of archive data are available at most RFCs. In the example presented here, a common period of only five years of precipitation forecasts and observations were available and only four years of temperature were available. Furthermore most RFCs do not have archived precipitation forecasts for day-2 and beyond. A related concern is that the procedures used to make precipitation and temperature forecasts are gradually changing. Especially in the case of atmospheric model forecasts, the local effects of changes might be much more dramatic than those that occur on average over the model domain. These could have very important local hydrological forecast implications. Another implication of having only limited archive data to estimate model parameters is that resulting ensemble forecasts may not be reliable unless

the forecast skill is very high. As the correlation parameter ρ decreases, the skill in the single-value forecast decreases and the forecast distribution should approach the long-range climatological distribution. But if only a few years of archive data are used to estimate model parameters the forecast distribution will approach the climatology of only those few years, not the long-range climatology. Also the single-value forecast used to estimate the conditional distribution of future events could be a deterministic forecast or a mean value of ensemble members. Therefore, techniques are needed to use other data such as archived atmospheric model ensemble forecast data, to develop improved parameter estimation procedures. One approach would be to infer parameter values using archived global ensemble re-forecast data sets from a fixed model, which are available for the entire USA from 1979. Future modifications to the proposed procedure may also be desirable, especially if there were an estimate of probability of precipitation available as well as a forecast of precipitation amount.

Further verification work is needed to use independent datasets of forecast and observed pairs for parameter estimation and verification. Also verification of ensemble forecasts is an important new topic. Some obvious verification statistics exist. However careful attention is needed to develop better statistics and to understand the properties of both existing and new statistics. In addition, the hydrology community needs to work together to agree on a few key statistics that normally would be included in any new study.

References

Bell, T. L.: A space-time stochastic model of rainfall for satellite remote-sensing studies, J. Geophys. Res., 92, 9631–9643, 1987.

Bellerby, T. L. and Sun, J.: Probabilistic and ensemble representations of the uncertainty in an IR/microwave satellite precipitation product, J. Hydrometeorol., 6(6), 1032–1044, 2005.

Carpenter, T. M. and Georgakakos, K. P.: Assessment of Folsom Lake response to historical and potential future climate scenarios: 1. Forecasting, J. Hydrol., 249, 148–175, 2001.

Clark, M. and Hay, L.: Use of medium-range weather forecasts to produce predictions of streamflow, J. Hydrometeorol., 5, 15–32, 2004.
Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake Shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields, J. Hydrometeorol., 5 , 243–262, 2004.
Day, G. N.: Extended streamflow forecasting using NWSRFS, ASCE J. Water Res. Plan. Manage., 111, 157–170, 1985.
Day, G. N.: A methodology for updating a conceptual snow model with snow measurements, Office of Hydrology, NOAA Technical Report NWS 43, Silver Spring, Maryland, 1990.
Georgakakos, K. P., Seo, D. J., Gupta, H., Schaake, J. C., and Butts, M. B.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, J. Hydrol., 298, 222–241, 2004.
Glahn, H. R. and Lowry, D. A.: The use of model output statistics (MOS) in objective weather forecasting, J. Appl. Meteorol., 11, 1203–1211, 1972.
Hamill, T., Whitaker, J. S., and Wei, X.: Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts, Mon. Wea. Rev., 132, 1434–1447, 2004.
Hirsch, R., Schaake, J. C., and Sheer, D .P.: Assessment of the current Occoquon water supply situation. ICPRB Publication Number: ICPRB-M-4, 1977.
Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecasting, 15, 559–570, 2000.
Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification, A Practitioner's Guide in Atmospheric Science, Wiley, 240 pp, 2003.
Kalnay, E., Kanamitsu, M., Kistler, R., et al.: The NCEP/NCAR 40-Year Reanalysis Project. Bull. Am. Meteorol. Soc., 77, 437–471, 1996.
Kelly, K. S. and Krzysztzyfowicz, R.: A bivariate meta-Gaussian density for use in hydrology, Stochastic Hydrol. Hydraul., 11, 17–31, 1997.
Kelly, K. S. and Krzysztzyfowicz, R.: Precipitation uncertainty processor for probabilistic river stage forecasting, Water Resour. Res., 35, 2643–2653, 2000.
Kistler, R., Kalnay, E., Collins, W., et al.: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, Bull. Am. Meteorol. Soc., 82, 247–267, 2001.
Krzysztzyfowicz, R.: Probabilistic hydrometeorological forecasts: Toward a new era in operational forecasting, Bull. Am. Meteorol. Soc., 79, 243–251, 1998.
Krzysztzyfowicz, R.: Integrator of uncertainties for probabilistic river stage forecasting:

precipitation-dependent model, J. Hydrol., 249, 69–85, 2001.
Krzysztzyfowicz, R. and Herr, H. D.: Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model, J. Hydrol., 249, 46–68, 2001.
Krzysztzyfowicz, R. and Kelly, K. S.: Hydrologic uncertainty processor for probabilistic river stage forecasting, Water Resour. Res., 36, 3265–3277, 2000.
NWSRFS: National Weather Service River Forecast System (NWSRFS) User Manual Documentation, VI.2.5. Extended Streamflow Prediction system (ESP) application information, National Weather Service documentation, Silver Spring, Maryland, 2000.
Smith, J. A., Day, G. N., and Kane, M. D.: Nonparametric framework for long-range streamflow forecasting, ASCE J. Water Res. Plan. Manage., 118, 82–91, 1992.
Schaake, J. C. and Larson, L.: Ensemble Streamflow Prediction (ESP). Preprints, Special Symposium on Hydrology, AMS, Boston MA, pp J19–J24, 1998.
Seo, D. J., Perica, S., Welles, E., and Schaake, J. C.: Simulation of precipitation fields from probabilistic quantitative precipitation forecast, J. Hydrol., 239, 203–229, 2000.
Tustison, B., Harris, D., and Fourfoula-Georgiou, E.: Scale issues in verification of precipitation forecasts, J. Geophys. Res., 106(D11), 775–778, 2001.
Weygandt, S., Loughe, A. F., Benjamin, S. G., and Mahoney, J. L.: Scale sensitivities in model precipitation skill scores during IHOP, 22nd Conference on Severe Local Storms, AMS, Hyannis, MA, 4–8 October, 2004.
Whitaker, J. S. and Loughe, A. F.. The relationship between ensemble spread and ensemble mean skill, Mon. Wea, Rev., 126, 3293–3302, 1998.
Wilks, D. S.: Statistical Methods in the Atmospheric Sciences: An Introduction, San Diego: Academic Press, 467 pp, 1995.

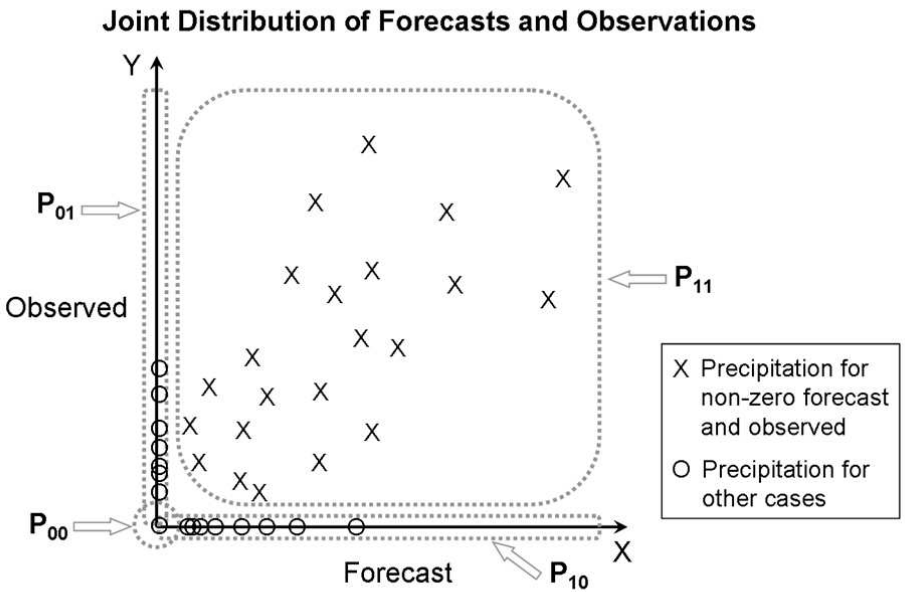


Fig. 1. Schematic diagram showing the four sub-regions of the joint distribution of precipitation forecasts and observations. Forecast is variable X, observation is variable Y. Symbols P_{00}, P_{01}, P_{10}, and P_{11} correspond to the fraction of forecast and observation pairs that lie in each of the corresponding sub-regions.

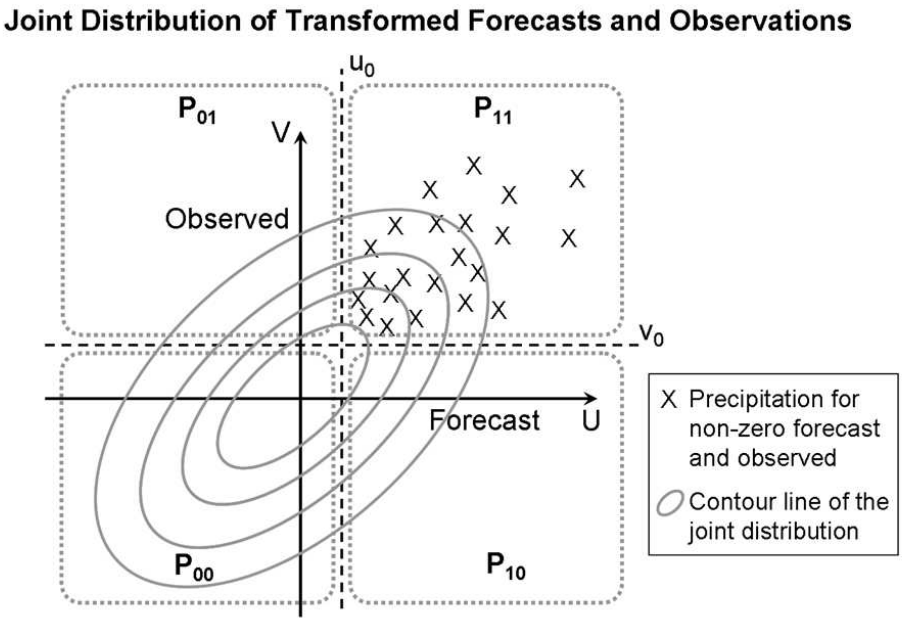


Fig. 2. Joint distribution of transformed values of precipitation forecasts and observations. Forecast variable X is transformed to variable U. Observed variable Y is transformed to variable V. The four sub-regions in Fig. 1 are shown in this figure. In normal space, only the P_{11} fraction of forecast and observed pairs can be represented with points. Value $U=u_0$ corresponds to value $X=0$. Value $V=v_0$ corresponds to value $Y=0$.

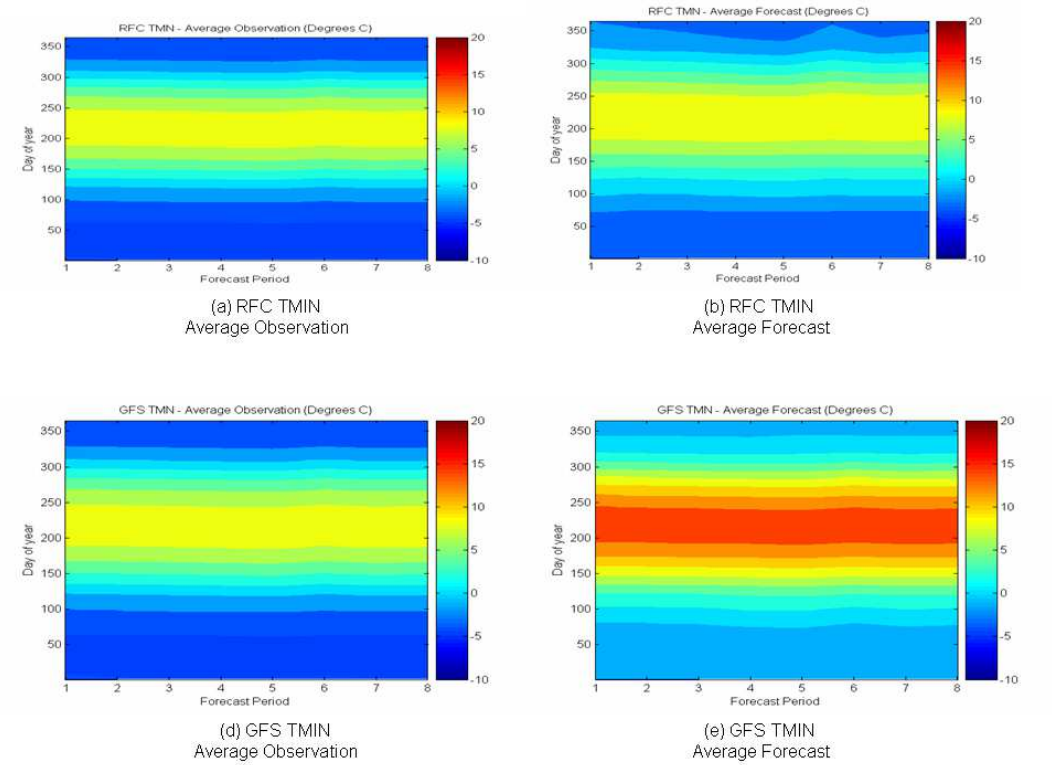


Fig. 3. Average Minimum Temperatures: **(a)** Average observation corresponding to RFC forecasts, **(b)** Average RFC Forecast, **(c)** Average observation corresponding to GFS forecasts, **(d)** Average GFS Forecast.

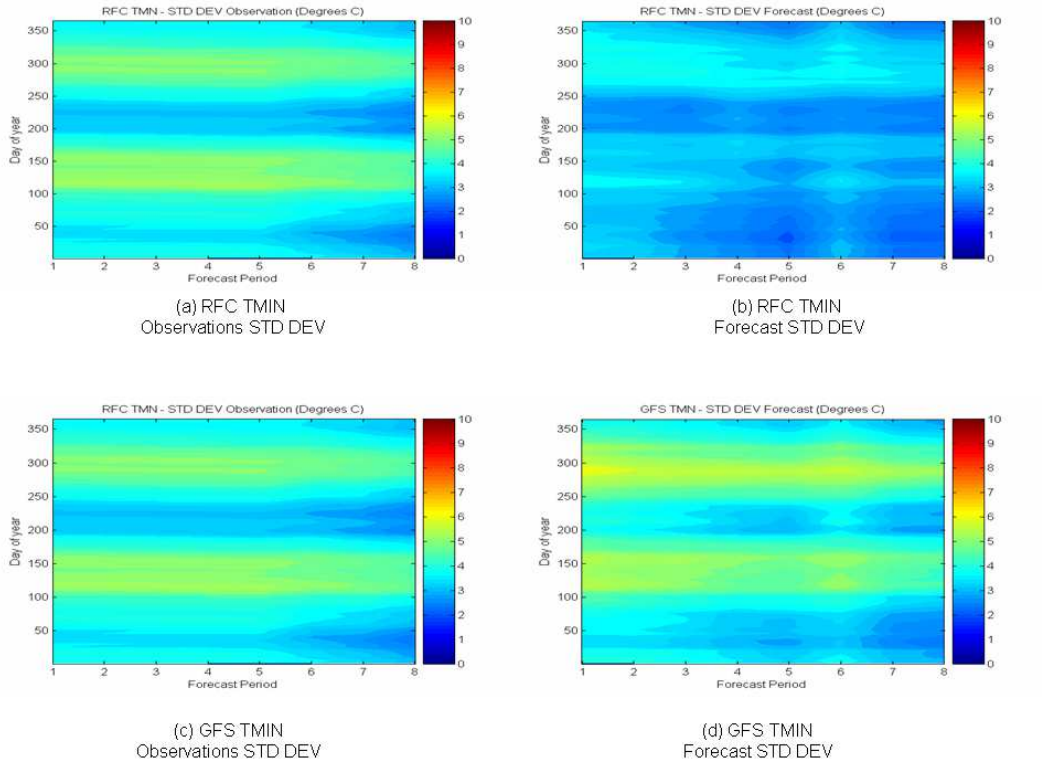


Fig. 4. Standard Deviation of Minimum Temperatures: **(a)** Standard deviation of observations corresponding to RFC forecasts, **(b)** Standard deviation of RFC Forecasts, **(c)** Standard deviation of observations corresponding to GFS forecasts, **(d)** Standard deviation of GFS Forecasts.

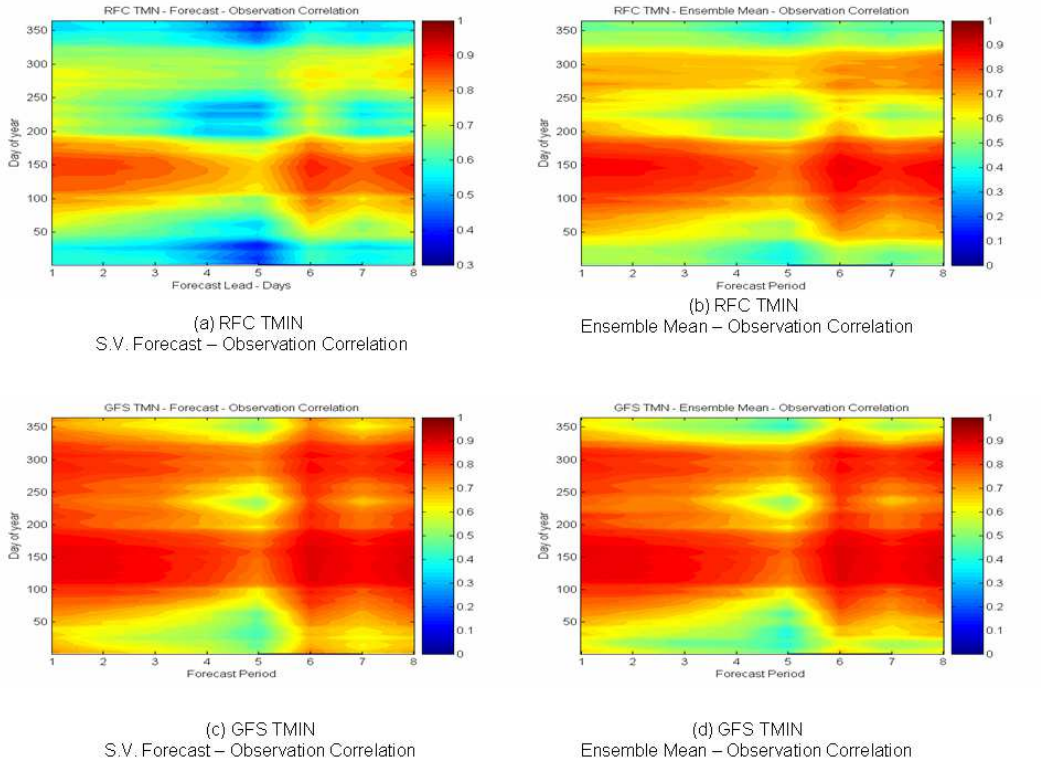


Fig. 5. Correlation Coefficients between Minimum Temperature Forecasts and Observations: **(a)** Single-value RFC forecasts, **(b)** Ensemble mean forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble mean forecasts based on GFS single-value forecasts.

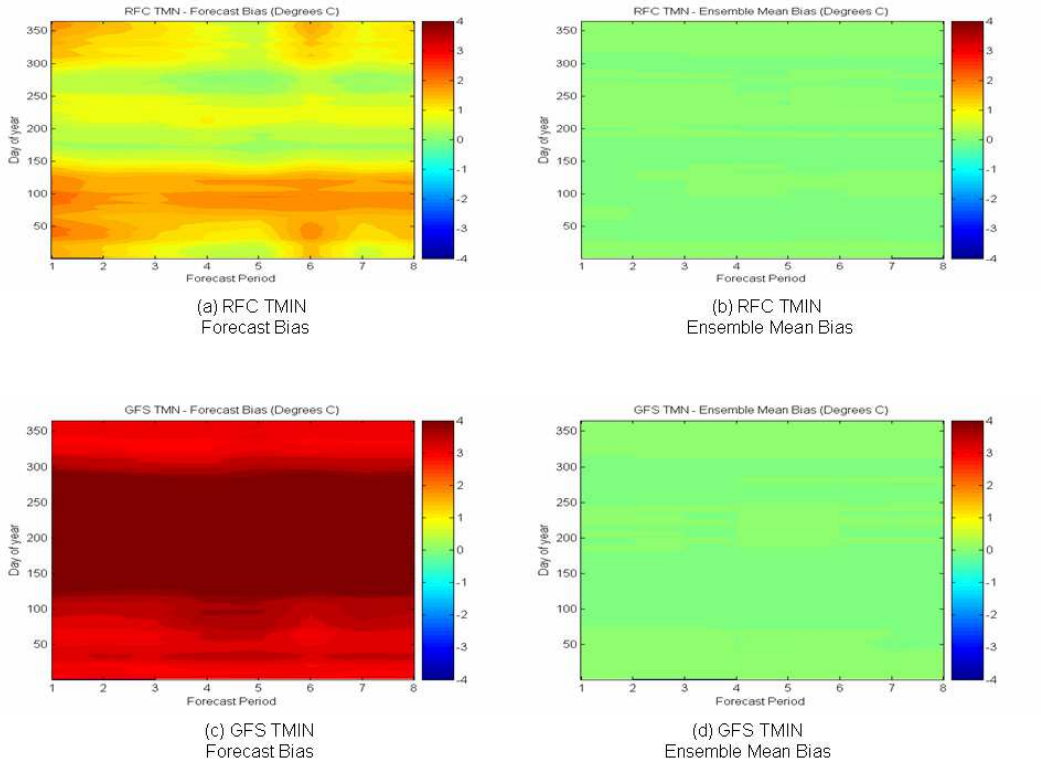
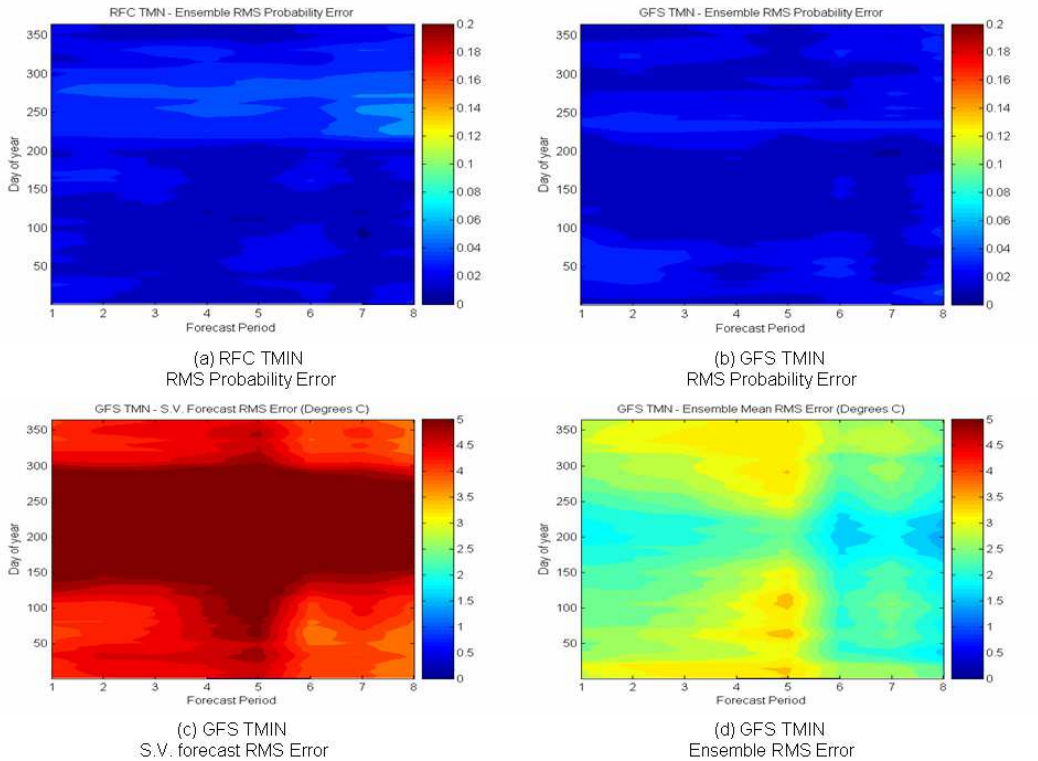


Fig. 6. Minimum Temperature Forecast Bias: **(a)** Single-value RFC forecasts, **(b)** Ensemble mean forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble mean forecasts based on GFS single-value forecasts.



(a) RFC TMIN
RMS Probability Error

(b) GFS TMIN
RMS Probability Error

(c) GFS TMIN
S.V. forecast RMS Error

(d) GFS TMIN
Ensemble RMS Error

Fig. 7. Minimum Temperature Forecast RMS Errors: **(a)** RMS error of single-value RFC forecasts, **(b)** RMS error ensemble mean forecasts based on single-value RFC forecasts, (b) RMS error of probability forecasts based on single-value GFS forecasts, **(c)** RMS error of single-value GFS forecasts, **(d)** RMS error ensemble mean forecasts based on single-value GFS forecasts.

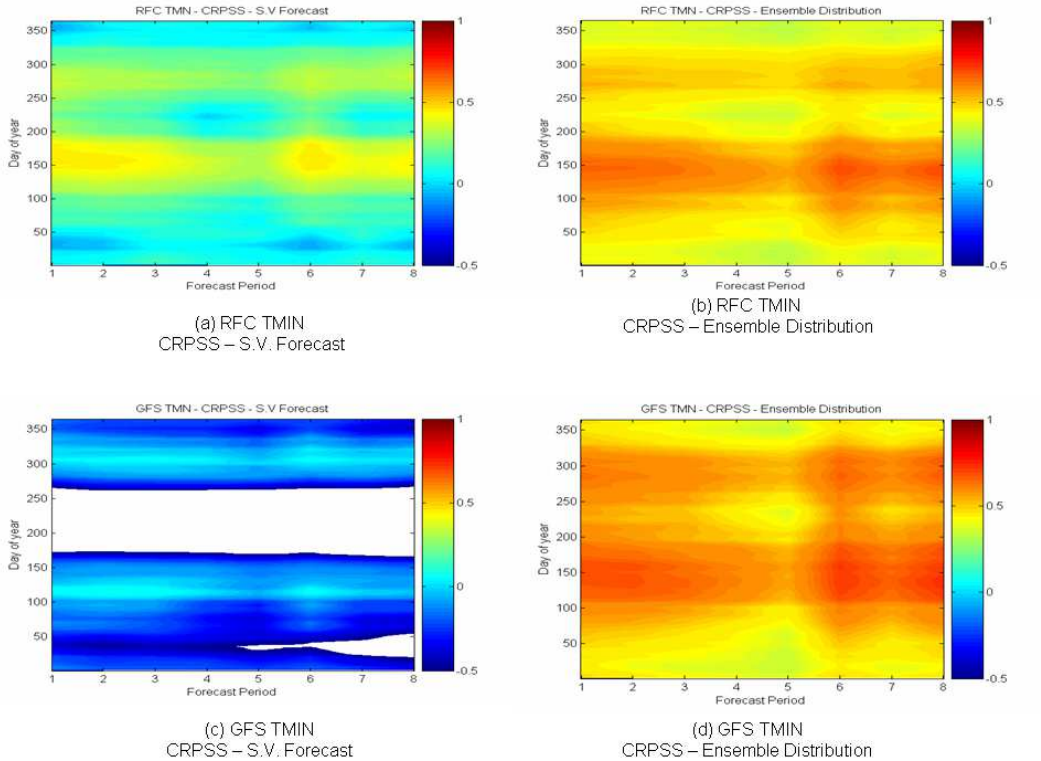
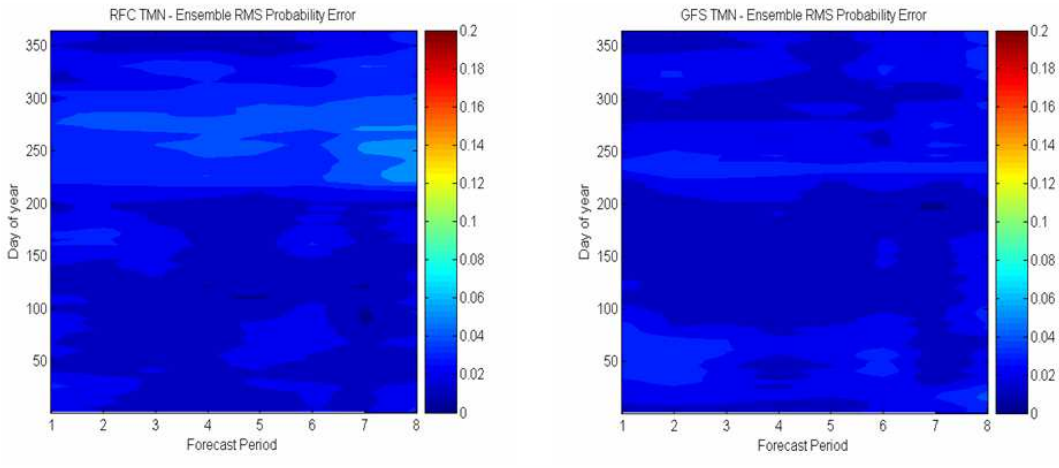


Fig. 8. Minimum Temperature Forecast Continuous Rank Probability Skill Score: **(a)** Single-value RFC forecasts, **(b)** Ensemble mean forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble mean forecasts based on GFS single-value forecasts.



(a) RFC TMIN
RMS Probability Error

(b) GFSFC TMIN
RMS Probability Error

Fig. 9. Minimum Temperature Forecast RMS Probability Errors: **(a)** RMS error of probability forecasts based on single-value RFC forecasts, **(b)** RMS error of probability forecasts based on single-value GFS forecasts.

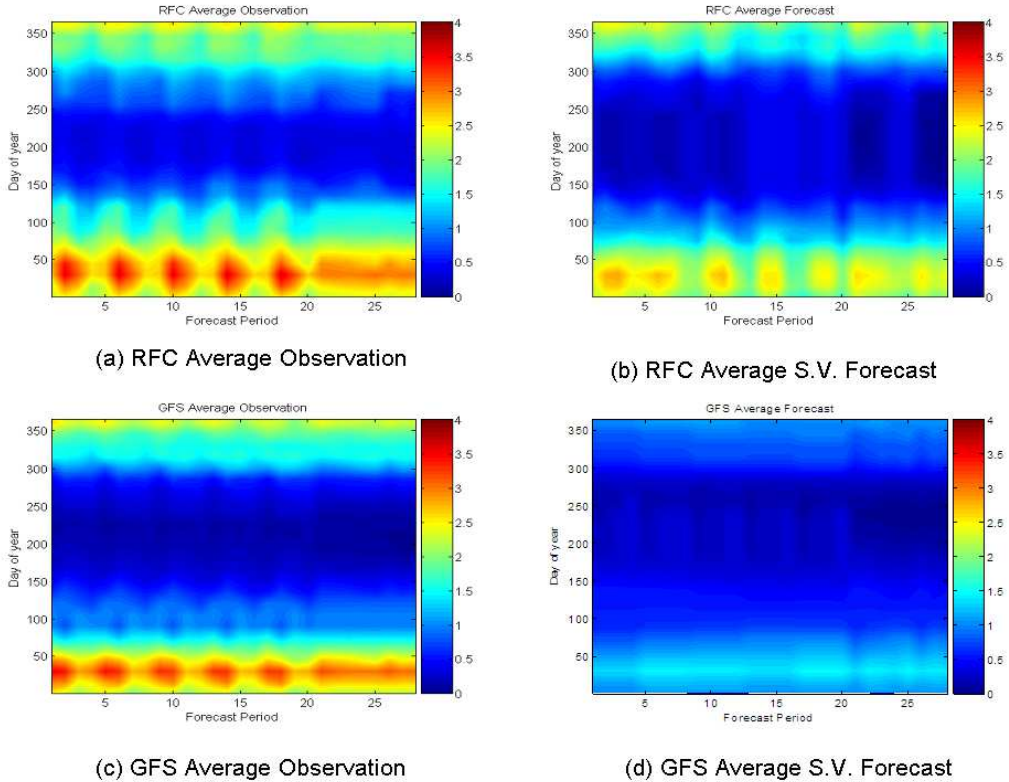


Fig. 10. Average Precipitation (mm/6 h): **(a)** Average observations corresponding to single-value RFC forecasts, **(b)** Average of single-value RFC forecasts, **(c)** Average of observations corresponding to single-value GFS forecasts, **(d)** Average of single-value GFS forecasts.

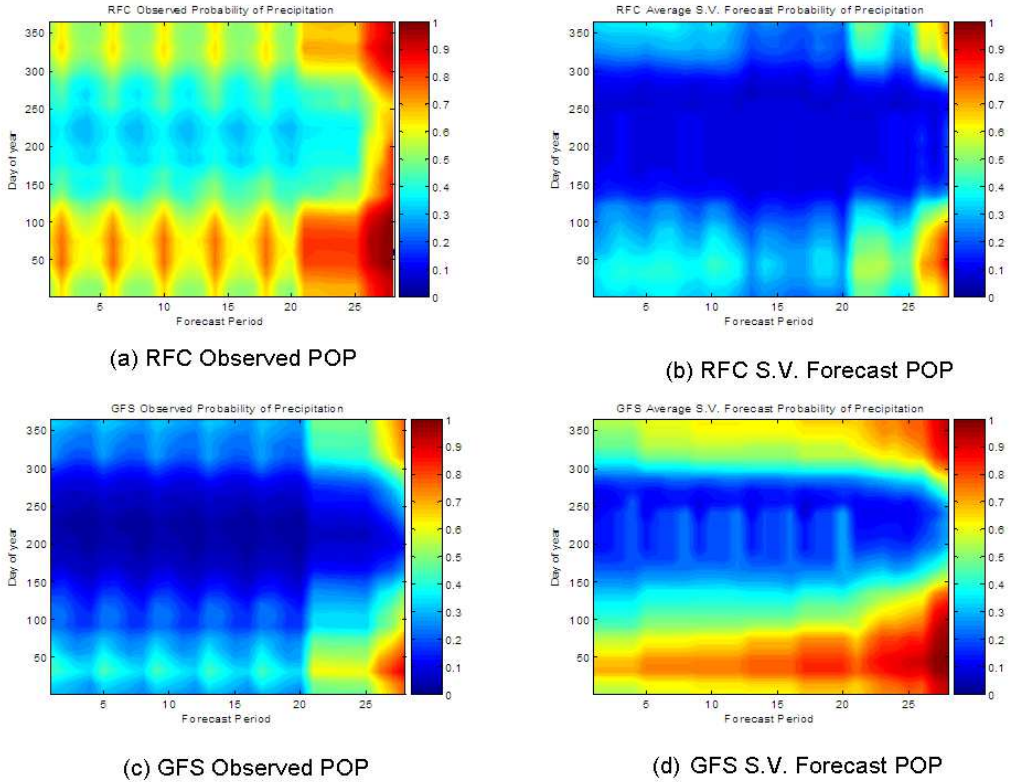
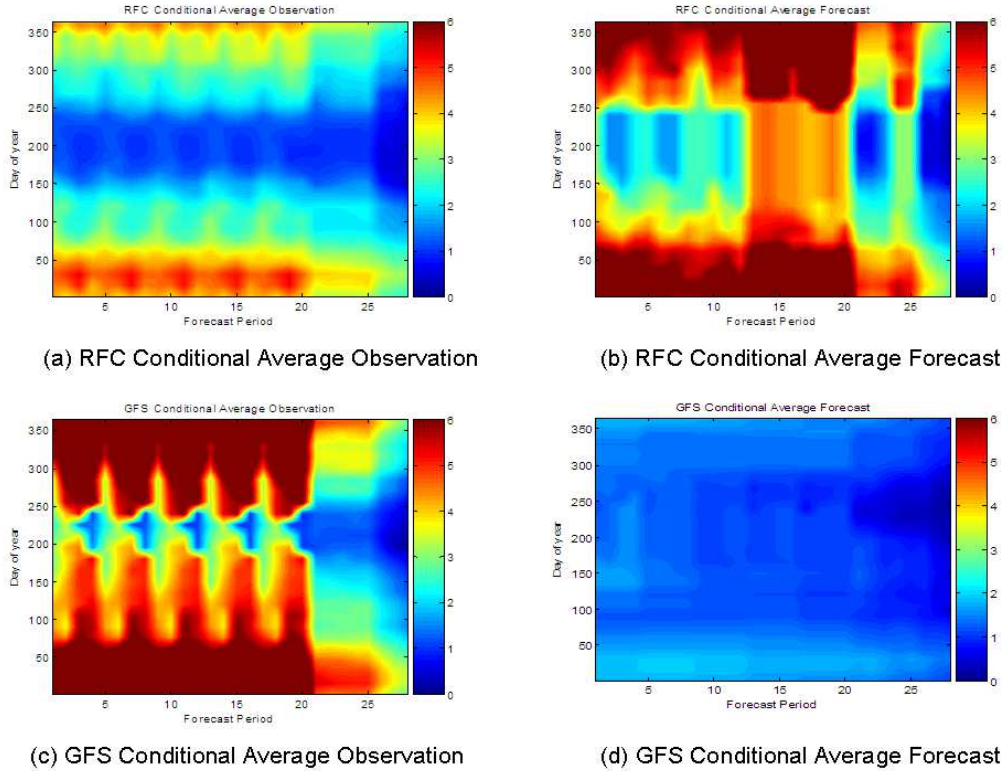


Fig. 11. Probability of Precipitation (POP): **(a)** POP corresponding to single-value RFC forecasts, **(b)** POP for single-value RFC forecasts, **(c)** POP for observations corresponding to single-value GFS forecasts, **(d)** POP for single-value GFS forecasts.



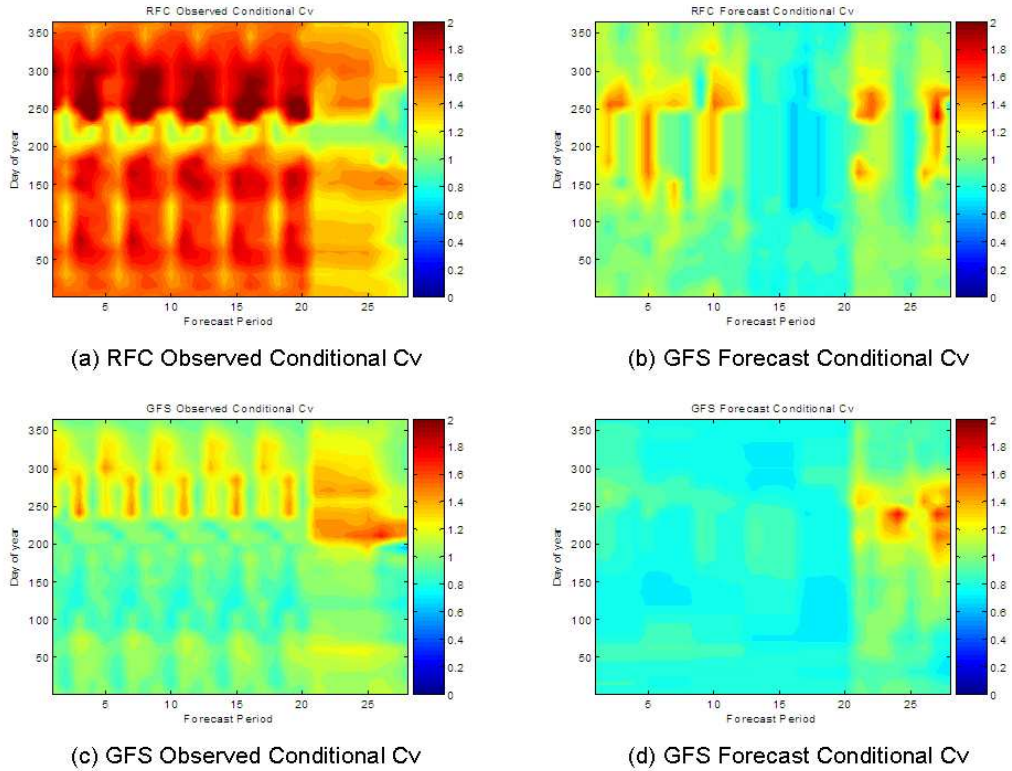
(a) RFC Conditional Average Observation

(b) RFC Conditional Average Forecast

(c) GFS Conditional Average Observation

(d) GFS Conditional Average Forecast

Fig. 12. Conditional Average Precipitation (during wet events) (mm/6 h) : **(a)** Conditional average observations corresponding to single-value RFC forecasts, **(b)** Conditional average of single-value RFC forecasts, **(c)** Conditional average of observations corresponding to single-value GFS forecasts, **(d)** Conditional average of single-value GFS forecasts.



(a) RFC Observed Conditional Cv

(b) GFS Forecast Conditional Cv

(c) GFS Observed Conditional Cv

(d) GFS Forecast Conditional Cv

Fig. 13. Conditional Coefficient of Variation of Precipitation (during wet events): **(a)** Conditional coefficient of variation of wet observations corresponding to single-value RFC forecasts, **(b)** Conditional coefficient of single-value RFC forecasts, **(c)** Conditional coefficient of variation of observations corresponding to single-value GFS forecasts, **(d)** Conditional coefficient of variation of single-value GFS forecasts.

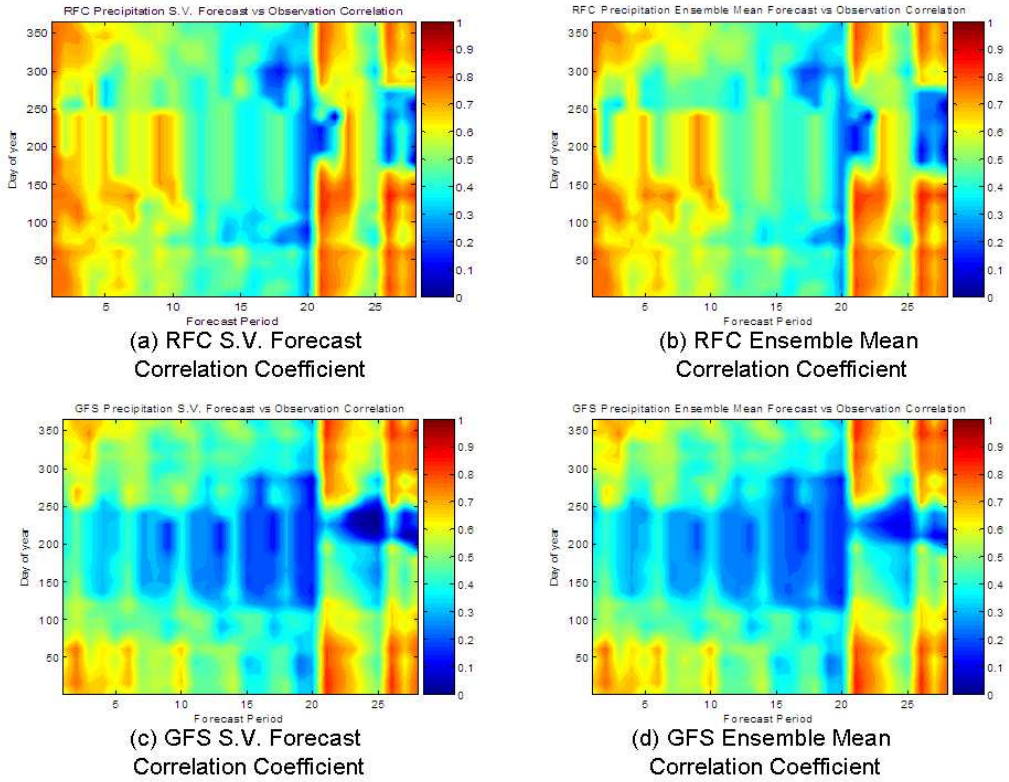


Fig. 14. Correlation Coefficients between Precipitation Forecasts and Observations: **(a)** Single-value RFC forecasts, **(b)** Ensemble mean forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble mean forecasts based on GFS single-value forecasts.

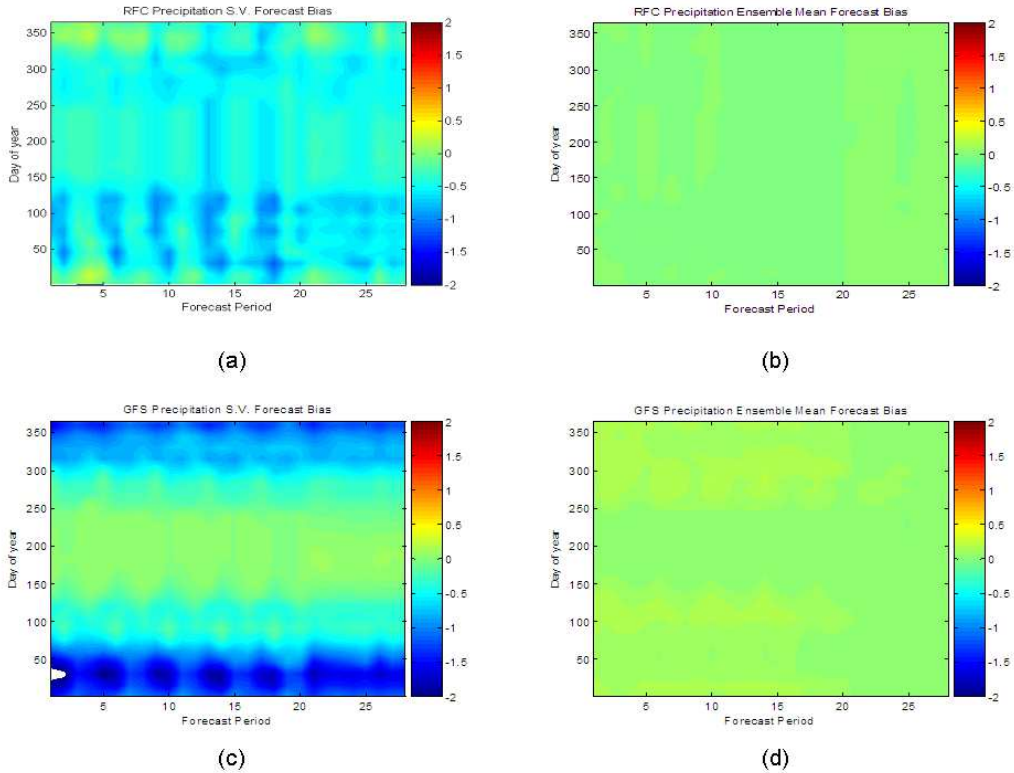


Fig. 15. Precipitation Forecast Bias (mm/6 h): **(a)** Single-value RFC forecasts, **(b)** Ensemble mean forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble mean forecasts based on GFS single-value forecasts.

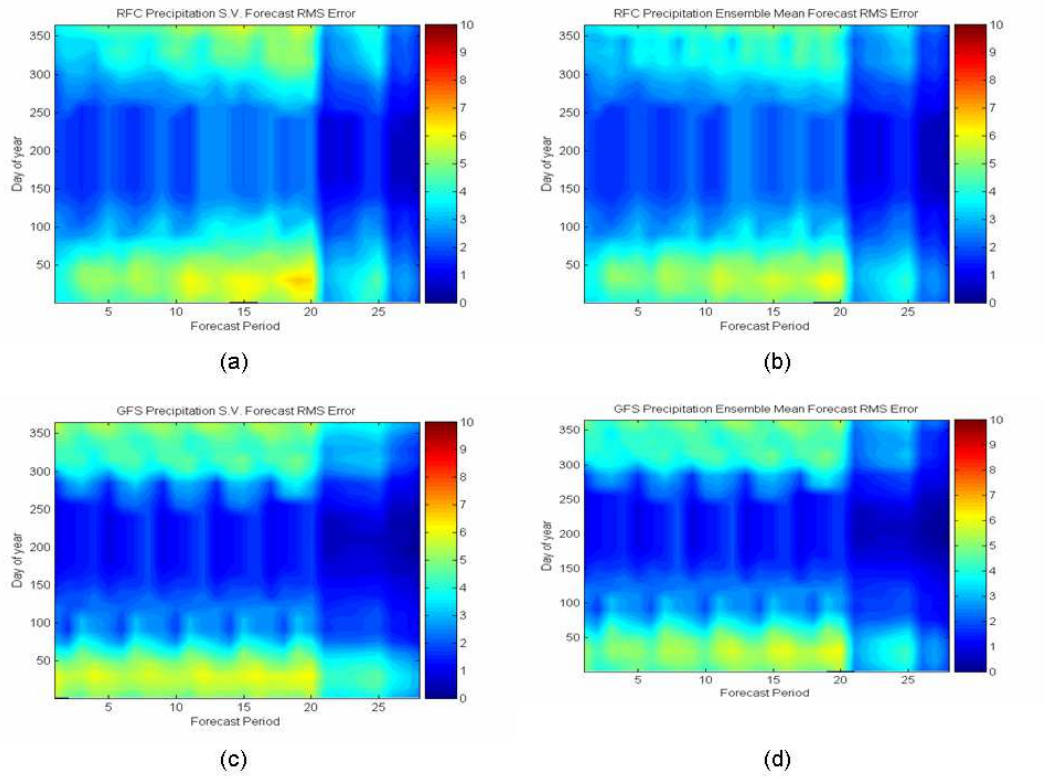


Fig. 16. Precipitation Forecast Bias (mm/6 h): **(a)** Single-value RFC forecasts, **(b)** Ensemble mean forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble mean forecasts based on GFS single-value forecasts.

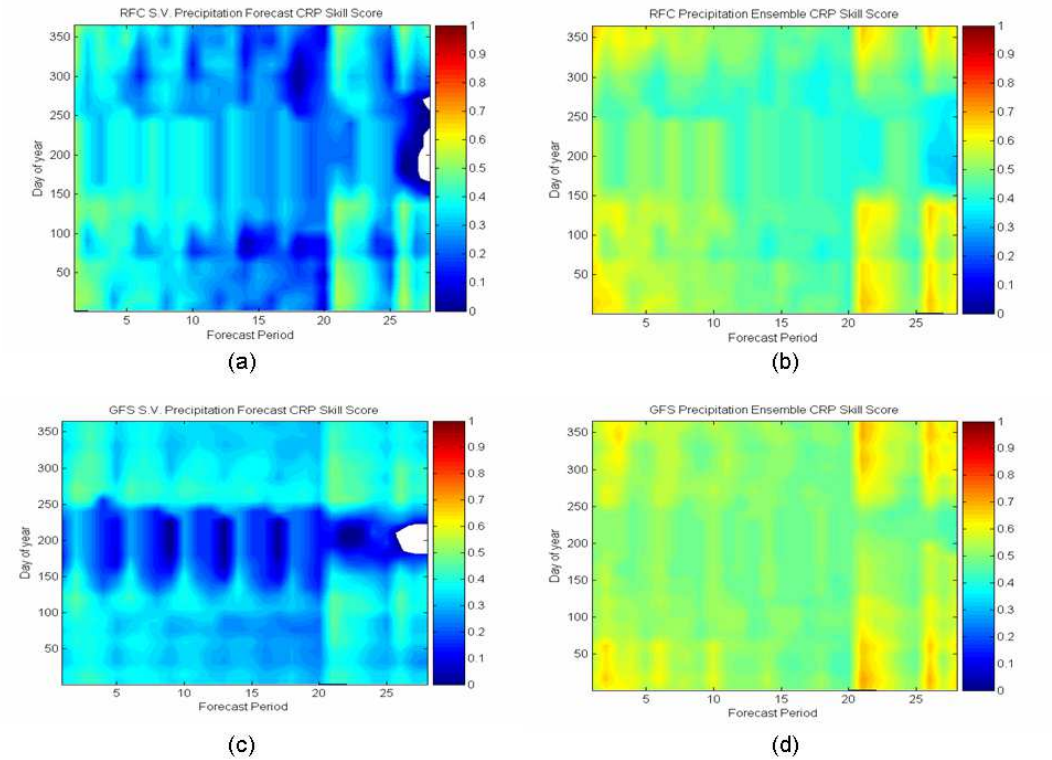


Fig. 17. Continuous Rank Probability Skill Score: **(a)** Single-value RFC forecasts, **(b)** Ensemble forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble forecasts based on GFS single-value forecasts.

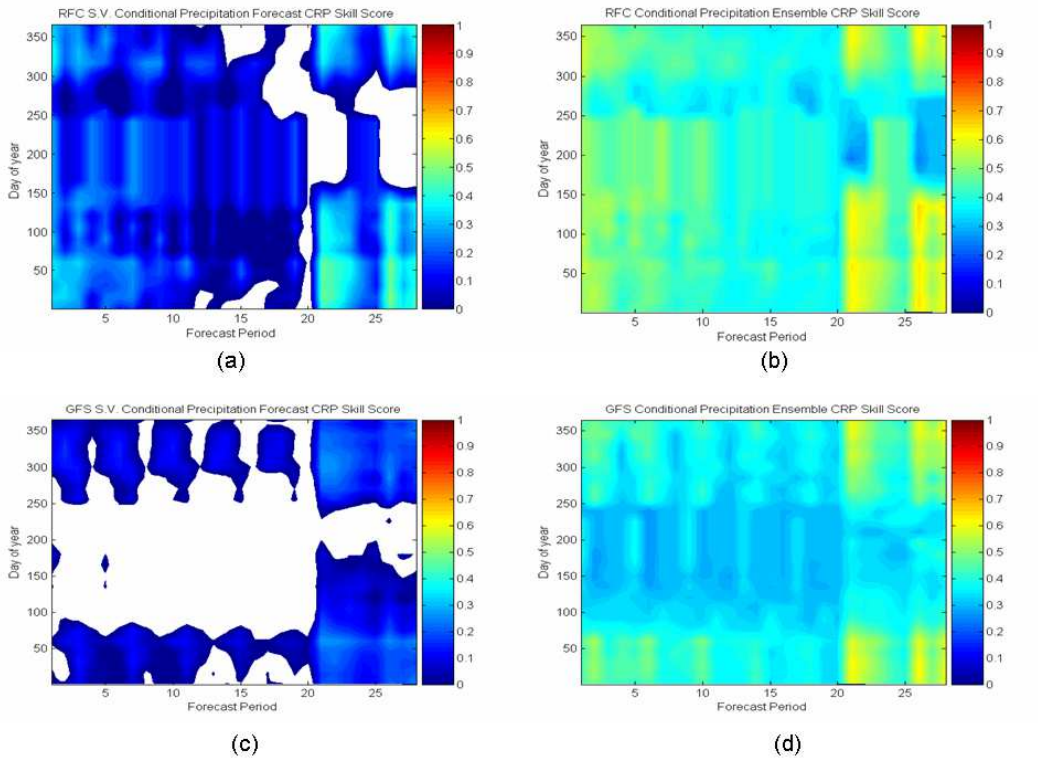


Fig. 18. Continuous Rank Probability Skill Score for Wet Events: **(a)** Single-value RFC forecasts, **(b)** Ensemble forecasts based on RFC single-value forecasts, **(c)** Single-value GFS forecasts, **(d)** Ensemble forecasts based on GFS single-value forecasts.

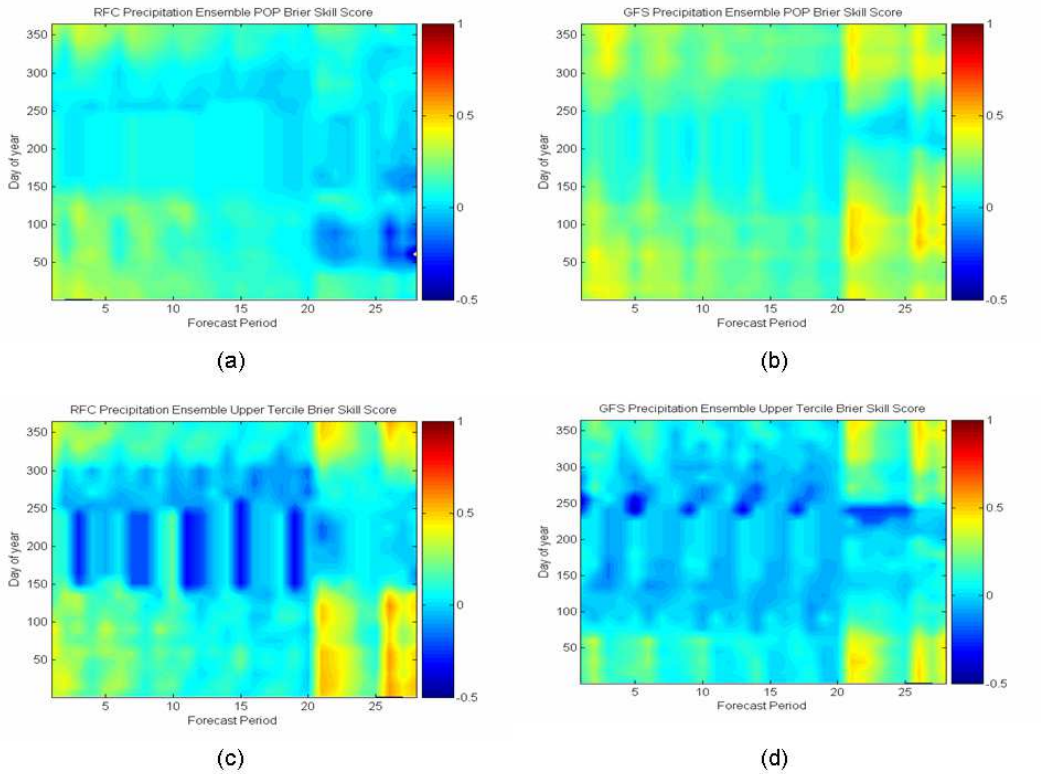


Fig. 19. Brier Skill Scores: **(a)** Probability of precipitation based on RFC forecasts, **(b)** Probability of precipitation based on GFS, **(c)** Probability of precipitation greater than upper tercile of conditional precipitation distribution using RFC forecasts, **(d)** Probability of precipitation greater than upper tercile of conditional precipitation distribution using GFS forecasts.

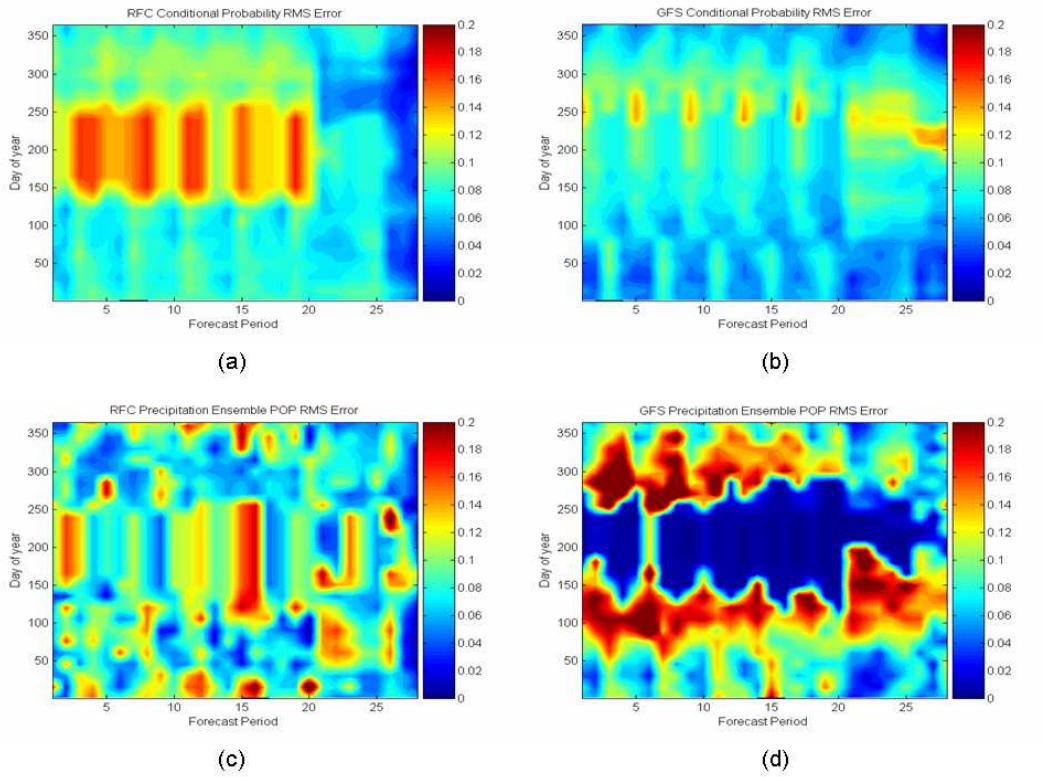


Fig. 20. RMS Probability Errors: **(a)** RFC based probability of precipitation amount during wet events, **(b)** GFS based probability of precipitation amount during wet events, **(c)** Probability of precipitation based on RFC forecasts, **(d)** Probability of precipitation based on GFS forecasts.