Uncertainty Propagation of Regional Climate Model Precipitation Forecasts to Hydrologic Impact Assessment

PHAEDON C. KYRIAKIDIS,* NORMAN L. MILLER, AND JINWON KIM

Regional Climate Center, Lawrence Berkeley National Laboratory, Berkeley, California

ABSTRACT

A Monte Carlo framework is adopted for propagating uncertainty in dynamically downscaled seasonal forecasts of area-averaged daily precipitation to associated streamflow response calculations. Daily precipitation is modeled as a mixture of two stochastic processes: a binary occurrence process and a continuous intensity process, both exhibiting serial correlation. The parameters of these processes (e.g., the proportion of wet days and the average wet-day precipitation intensity in a month) are derived from the forecast record. Parameter uncertainty is characterized via an empirical Bayesian model, whereby such parameters are modeled as random with a specific joint probability distribution. The hyperparameters specifying this probability distribution are derived from historical precipitation records at the study basin. Simulated parameter values are then generated using the Bayesian model, leading to alternative synthetic daily precipitation records simulated via the stochastic precipitation model. The set of such synthetic precipitation records is finally input to a physically based deterministic hydrologic model for propagating uncertainty in forecasted precipitation to hydrologic impact assessment studies.

The stochastic simulation approach is applied for generating an ensemble (set) of synthetic area-averaged daily precipitation records at the Hopland basin in the northern California Coast Range for the winter months (December through February: DJF) of 1997/98. The parameters of the stochastic precipitation model are derived from a seasonal precipitation forecast based on the Regional Climate System Model (RCSM), available at a 36-km^2 grid spacing. The large-scale forcing input to RCSM for dynamical downscaling was a seasonal prediction of the University of California, Los Angeles, Atmospheric General Circulation Model. A semidistributed deterministic hydrologic model ("TOPMODEL") is then used for calculating the streamflow response for each member of the area-averaged precipitation ensemble set. Uncertainty in the parameters of the stochastic precipitation model is finally propagated to associated streamflow response, by considering parameter values derived from historical (DJF 1958–92) area-averaged precipitation records at Hopland.

1. Introduction

The use of downscaled climate projections has become increasingly important for hydrologic impact assessment studies, such as streamflow prediction at seasonal or interannual timescales (Leung et al. 1996, 1999; Miller and Kim 1996; Kim et al. 1998, 2000; Miller et al. 1999). One of the critical tasks in hydrologic forecasting is predicting the probability that extreme events, for example, droughts or floods, will occur during the forecast period. Accurate prediction of such hydrologic events, however, is extremely difficult, because their occurrence over complex terrain depends not only on

the amount of precipitation, but also on its temporal sequence, both affecting feedbacks in the form of antecedent soil moisture conditions. Understanding the mechanisms and interactions involved in such a complex hydroclimatic system poses a challenging fundamental research theme in hydrology (Entekhabi et al. 1999).

Flood forecasting utilizing deterministic hydrologic models requires input precipitation forcing on daily, or shorter, timescales. Daily sequences of precipitation events from regional climate forecasts, however, are not as accurate as their average values over monthly or seasonal timescales (Kim et al. 2000). In addition, uncertainties in the parameterization and accurate description of variables affecting predictions of global atmospheric models, which in turn drive any dynamical downscaling procedure as initial and time-varying boundary conditions, propagate to regional climate forecasts at monthly or seasonal timescales. Consequently, there is a need for modeling uncertainty in projected daily precipitation records and evaluating its impact to hydrologic studies (e.g., Lettenmaier 1994).

* Current affiliation: Department of Geography, University of California, Santa Barbara, Santa Barbara, California.

Corresponding author address: Phaedon C. Kyriakidis, Dept. of Geography, University of California, Santa Barbara, Ellison Hall 3611, Santa Barbara, CA 93106-4060.
E-mail: phaedon@geog.ucsb.edu

It should be stressed that any uncertainty assessment is necessarily subjective: one almost always decides to "freeze" specific sources of uncertainty affecting the problem at hand. In a hydrologic modeling context, for example, one might decide to adopt a semidistributed framework instead of a fully distributed one. Consequently, projected uncertainty intervals are specific to the model structure (or modeling framework) used to evaluate them. In the context of uncertainty assessment in streamflow predictions, one of the most dominant components of uncertainty is that of input precipitation forcing. In this paper, we focus on such precipitation-induced uncertainty, thus ignoring uncertainty in other hydrologic variables, for example, parameter uncertainty in deterministic hydrologic models and/or measurement errors. In particular, we study the impact of area-averaged precipitation forcing to associated streamflow response, thus also ignoring the effects of precipitation spatial variability. A spatially explicit modeling framework, which accounts for precipitation spatial variability, is currently in development and will be reported in the near future.

In this paper, we use "TOPMODEL" (Beven and Kirkby 1979), a semidistributed rainfall–runoff model, for evaluating streamflow response to input precipitation forcing. TOPMODEL may be viewed as a system of vertical reservoirs (interception zone storage, unsaturated infiltration zone storage, and saturated zone storage) controlled by Darcy's law and water mass continuity. Soil moisture accounting and lateral transport in TOPMODEL are based on two assumptions: 1) the saturated zone dynamics are approximated by successive steady-state representations and 2) the hydraulic gradient of the saturated zone is parallel to the surface topography. In the current study, we use the semidistributed version of TOPMODEL with topographic parameters derived from 50-m digital elevation model data. The advantage of this model is the low number of parameters required for optimization while maintaining an adequate physical process description. The primary disadvantages are the above two assumptions, which are not always valid.

In a Monte Carlo simulation context, multiple alternative simulated realizations of daily precipitation can be processed by a deterministic rainfall–runoff model, for example, to yield multiple forecasts of riverflow for uncertainty characterization and hydrologic design (Salas 1993; Krzysztofowicz 1998). The critical task is to ensure that both synthetic and observed precipitation records exhibit similar characteristics, such as serial correlation (persistence) and seasonality in temporal variability. In the context of stochastic temporal downscaling of monthly or seasonal climate projections, an additional constraint is that of reproduction of statistics on the larger levels of aggregation (Wilby et al. 1998).

For long-term (seasonal) streamflow forecasting, we use a dynamically downscaled seasonal forecast of area-averaged daily precipitation (Kim et al. 2000) to infer parameters of a stochastic precipitation model. From this stochastic model, we generate a set of alternative synthetic area-averaged daily precipitation records, which share certain characteristics with the forecast record. Such characteristics include serial autocorrelation in both daily precipitation intermittence and intensity, as well as monthly statistics derived from that record, such as wet-day proportions and average wet-day precipitation intensity values in any month. The synthetic precipitation records are subsequently input to TOPMODEL for generating the corresponding streamflow realizations. The set of these alternative, TOPMODEL-derived, synthetic streamflow records constitutes a model of uncertainty regarding important streamflow characteristics, for example, flood probabilities. Such a Monte Carlo procedure allows propagation of uncertainty in the forecasted precipitation to associated hydrologic predictions.

The parameters of the stochastic precipitation model, however, are themselves uncertain given that they are derived from an uncertain forecasted precipitation record. We adopt a Bayesian framework (Box and Tiao 1973; Gelman et al. 1995) for characterizing such parameter uncertainty and propagating it to associated streamflow calculations. For a Bayesian treatment of most uncertainty components inherent in streamflow forecasts the reader is referred to Krzysztofowicz (1999); classical applications of Bayesian analysis in stochastic hydrology can be found in Valdés et al. (1977) and Kitanidis (1986). In this paper, we develop an empirical Bayesian model (Carlin and Louis 2000), that is, an approximation to a full hierarchical Bayesian model (Gelman et al. 1995), whereby precipitation values for the forecast period are viewed as outcomes of the stochastic precipitation model, conditional to a set of parameter values derived from a dynamically downscaled precipitation forecast. Climatological parameter uncertainty (prior to any forecast information) is modeled using historical precipitation records at the study basin; the Bayesian paradigm is then adopted for updating this prior uncertainty model to account for the forecast parameter set.

Section 2 briefly presents the stochastic precipitation model used in this work. In section 3, the empirical Bayesian model for characterizing uncertainty in the parameters of the stochastic precipitation model is developed. In section 4, a case study is undertaken: a dynamically downscaled seasonal forecast of (area averaged) daily precipitation for the Hopland basin in the northern California Coast Range is first used to infer the parameters of the stochastic precipitation model. Parameter uncertainty is then characterized via the Bayesian model, using as prior information a set of historical parameters derived from historical area-averaged precipitation records for the same basin. Uncertainty propagation to streamflow response is achieved by first generating a set of alternative synthetic precipitation records and then using TOPMODEL to calculate the associated

streamflow for each synthetic record. Last, in section 5, some conclusions regarding the proposed uncertainty propagation procedure and its application to hydrologic impact assessments are drawn.

2. Stochastic daily precipitation model

Stochastic models for characterizing daily precipitation compose a dominant component of the literature of stochastic hydrology (Bras and Rodríguez-Iturbe 1985; Woolhiser 1992; Foufoula-Georgiou and Krajewski 1995). In this paper, we focus on mixture processes, one of the most frequently used stochastic frameworks for modeling daily precipitation (Woolhiser 1992; Salas 1993; Wilks 1995). Within the framework of mixture processes, precipitation modeling is decomposed into two separate tasks: (a) modeling of precipitation intermittence (occurrence/nonoccurrence of wet days) and (b) modeling of precipitation intensity (nonzero precipitation amounts on wet days). The model presented hereinafter is a generalization of the mixture models proposed by Chebaane et al. (1995), and Katz and Parlange (1995).

Let t_k denote a time step (day) in a forecast period composed of K days $\{t_k, k = 1, \ldots, K\}$. Under mixture models, the true unknown precipitation levels $\{z(t_k), k = 1, \ldots, K\}$ at the respective time steps are modeled as outcomes of a stochastic process:

$$\{Z(t), t \in K\} = \{I(t)X(t), t \in K\}, \qquad (1)$$

where $\{I(t), t \in K\}$ is a binary (two state) stochastic process modeling the temporal intermittence of precipitation, $\{X(t), t \in K\}$ is a continuous stochastic process modeling precipitation intensity, and the two processes are uncorrelated: $\text{Cov}\{I(t), X(t')\} = 0$, for every t and t'. For simplicity, the dependence of all processes on the spatial coordinates is dropped from the notation.

The objective is to generate a set (ensemble) of S alternative synthetic realizations of daily precipitation $\{z^{(s)}(t_k), k = 1, \ldots, K\}, s = 1, \ldots, S$, from the stochastic process $\{Z(t), t \in K\}$, where s denotes the sth member of the ensemble. This task is decomposed to the generation of S synthetic realizations of precipitation occurrence $\{i^{(s)}(t_k), k = 1, \ldots, K\}, s = 1, \ldots, S$, and precipitation intensity $\{x^{(s)}(t_k), k = 1, \ldots, K\}, s = 1, \ldots, S$. These realizations must share important characteristics of the observed precipitation records, such as statistics at the monthly level of aggregation (Wilks 1992) and daily persistence in both precipitation occurrence and intensity.

a. Precipitation occurrence

Precipitation occurrences $\{i(t_k), k = 1, \ldots, K\}$ over the time span of K days are modeled as outcomes of a binary stochastic process $\{I(t), t \in K\}$, that is, a collection of serially correlated random variables (RV) $I(t)$ defined as

$$I(t) = \begin{cases} 1 & \text{if } Z(t) > z_{\min}, \quad \forall t \in K \\ 0 & \text{if not,} \end{cases}$$

where $z_{\min}$ denotes a minimum precipitation threshold value, below which a day is considered as dry, for example, 0.1 mm day^{-1}.

Let $\mathbf{i}^{(s)} = \{i^{(s)}(t_n), n = 1, \ldots, N\}$ denote the set of N occurrence values simulated at N randomly visited time steps $\{t_n, n = 1, \ldots, N\}$, which have been visited prior to visiting step t_k (with $0 \le N < K$). A realization $\{i^{(s)}(t_k), k = 1, \ldots, K\}$ of the binary precipitation occurrence process $\{I(t), t \in K\}$ can be generated sequentially (Chang et al. 1984; Ripley 1987; Journel 1989) as

$$i^{(s)}(t_k) = j\{t_k; p^*[t_k \mid \mathbf{i}^{(s)}]\} \qquad \forall t_k \in K, \qquad (2)$$

where $j\{t_k; p^*[t_k \mid \mathbf{i}^{(s)}]\}$ is the indicator function of a random number $u^{(s)}(t_k)$ uniformly distributed in $[0, 1]$, defined as $j\{t_k; p^*[t_k \mid \mathbf{i}^{(s)}]\} = 1$ if $u^{(s)}(t_k) > p^*[t_k \mid \mathbf{i}^{(s)}]$, zero if not.

Term $p^*[t_k \mid \mathbf{i}^{(s)}]$ in Eq. (2) denotes an estimate of the unknown conditional probability density function (cpdf) $p[t_k \mid \mathbf{i}^{(s)}]$ of precipitation occurrence at any time step t_k, given the N previously simulated occurrence outcomes $\mathbf{i}^{(s)}$:

$$p[t_k \mid \mathbf{i}^{(s)}] = \text{Prob}\{I(t_k) = 1 \mid \mathbf{i}^{(s)}\} = E\{I(t_k) \mid \mathbf{i}^{(s)}\}, \quad (3)$$

which is a Bernoulli distribution with parameter $p[t_k \mid \mathbf{i}^{(s)}]$.

The least squares estimate $p^*[t_k \mid \mathbf{i}^{(s)}]$ of the unknown cpdf $p[t_k \mid \mathbf{i}^{(s)}]$ is written as

$$p^*[t_k \mid \mathbf{i}^{(s)}] = p_{c_k}(t_k) + \sum_{n=1}^{N} \nu_n(k)[i^{(s)}(t_n) - p_{c_n}(t_n)], \quad (4)$$

where $p_{c_k}(t_k)$ is the proportion of wet days of month c_k in which day t_k belongs.

Evidently, $p_{c_k}(t_k)$ need not be equal to $p_{c_n}(t_n)$; this allows consideration of nonstationary (monthly varying) wet-day proportions over the period of interest. In northern California, for example, the proportion of wet days in summer months (June, July, and August: JJA) is expected to be smaller than that in winter months (December, January, and February: DJF). Such nonstationary wet-day proportions within any given month can be derived from climate projections available for the forecast period $\{t_k, k = 1, \ldots, K\}$.

The N weights $\nu_n(k)$ in Eq. (4) are determined per solution of a system of normal equations [simple kriging (SK) system]:

$$\sum_{n'=1}^{N} \nu_{n'}(k) C_{R_I}(t_n - t_{n'}) = C_{R_I}(t_n - t_k),$$

$$n = 1, \ldots, N, \quad (5)$$

where $C_{R_I}(\tau = t - t') = E\{[I(t) - p_c(t)][I(t') - p_{c'}(t')]\}$ denotes the covariance of the residual random process $\{R_I(t) = [I(t) - p_c(t)], t \in K\}$.

In practice, covariance values $C_{R_I}(\tau)$ for various lag intervals τ are modeled via a parametric function $C_{R_I}(\tau; \delta)$, where δ denotes the correlation period (characteristic time) of the covariance model. In this paper, we assume for simplicity that the parametric covariance model $C_{R_I}(\tau; \delta)$ is defined by its functional form, for example, exponential, and a single parameter δ; more complex models can also be defined.

The system in Eq. (5) is a version of the Yule–Walker equations, widely used in hydrologic time series modeling [for example, Bras and Rodríguez-Iturbe (1985)]. The procedure of Eq. (4) can be seen as that of imposing a prior likelihood of precipitation occurrence at each time step t_k. Such prior likelihood acts in conjunction with the residual serial correlation modeled by $C_{R_I}(\tau; \delta)$. For example, if in a wet month [i.e., a month with relatively high prior probability $p_{c_k}(t_k)$] a precipitation event is simulated at day t_k, then the next day t_{k+1} is likely to be simulated as wet; this result is due to the high prior probability $p_{c_k}(t_k)$ and the serial correlation $C_{R_I}(\tau; \delta)$ quantifying the memory of the residual process.

It can be shown (Journel 1999) that synthetic realizations from this process reproduce the imposed autocovariance model $C_{R_I}(\tau; \delta)$ no matter its type, for example, exponential or spherical. This reproduction of a wider class of covariance models comes from the fact that N previously simulated values $\mathbf{i}^{(s)}$ are used for determining the cpdf in Eq. (4). The procedure in Eq. (2) can be viewed as an Nth-order discrete autoregressive representation with a random order for visiting the K time steps, or as an approximation of higher-order two-state Markov chains.

In all rigor, a synthetic precipitation occurrence record $\{i^{(s)}(t_k), k = 1, \ldots, K\}$ generated via the procedure described in this section should be denoted as $\{i^{(s)}(t_k), k = 1, \ldots, K; \mathbf{a}\}$, where $\mathbf{a}$ is a $(L_I \times 1)$ vector of parameter values $\mathbf{a} = [a_1, \ldots, a_{L_I}]'$ characterizing the binary occurrence process $\{I(t), t \in K\}$ (note that superscript $'$ denotes transposition). For the case of a season-long forecast period, $L_I = 4$, there are three entries in $\mathbf{a}$, one for each monthly proportion of wet days $p_{c_k}(t_k)$, and one last entry for the correlation period δ. For simplicity, the correlation period δ is considered to be constant within the season.

b. Precipitation intensity

Precipitation intensity is defined as the (positive) precipitation amount in a wet day t_k. Precipitation intensity outcomes $\{x(t_k), k = 1, \ldots, K\}$ over the time span of K days are modeled as outcomes of a continuous stochastic process $\{X(t), t \in K\}$, that is, a collection of serially correlated RVs $X(t)$ defined as

$$X(t) = \begin{cases} z(t) & \text{if } Z(t) > z_{\min}, \quad \forall t \in K \\ \text{undefined} & \text{if not.} \end{cases}$$

Let $\mathbf{x}^{(s)} = \{x^{(s)}(t_n), n = 1, \ldots, N\}$ denote the set of N intensity values simulated at N *randomly* visited time steps $\{t_n, n = 1, \ldots, N\}$, visited prior to visiting step t_k (with $0 \leq N < K$). A realization $\{x^{(s)}(t_k), k = 1, \ldots, K\}$ of the intensity process $\{X(t), t \in K\}$ can be generated as

$$x^{(s)}(t_k) = m^*[t_k | \mathbf{x}^{(s)}] + \sigma^*[t_k | \mathbf{x}^{(s)}] w^{(s)}(t_k),$$
$$\forall t_k \in K, \qquad (6)$$

where each term in the above equation is defined as follows.

Term $m^*[t_k | \mathbf{x}^{(s)}]$ of Eq. (6) denotes an estimate of the conditional mean of the RV $X(t_k)$ given the vector $\mathbf{x}^{(s)}$ of N previously simulated intensity values for the sth realization, that is, $m^*[t_k | \mathbf{x}^{(s)}] \cong E\{X(t_k) | \mathbf{x}^{(s)}\}$. This estimate is expressed as a weighted linear combination of the N previously simulated values,

$$m^*[t_k | \mathbf{x}^{(s)}] = m_{c_k}(t_k) + \sum_{n=1}^{N} \xi_n(k)[x^{(s)}(t_n) - m_{c_n}(t_n)], \quad (7)$$

where $m_{c_k}(t_k)$ denotes the average precipitation intensity of month c_k in which time step t_k belongs.

The N weights $\xi_n(k)$ of Eq. (7) are determined per solution of a system of normal equations (SK system):

$$\sum_{n'=1}^{N} \xi_{n'}(k) C_R(t_n - t_{n'}) = C_R(t_n - t_k),$$
$$n = 1, \ldots, N, \quad (8)$$

with $C_R(\tau) = E\{R(t)R(t + \tau)\}$ being the covariance of the zero-mean, unit-variance, residual process $\{R(t) = [X(t) - m(t)]/s_c(t), t \in K\}$, where $s_c(t)$ denotes the standard deviation of the precipitation intensity for month c to which time step t belongs.

In practice, covariance values $C_R(\tau)$ for various lag intervals τ are modeled via a parametric function $C_R(\tau; \eta)$, where η denotes the correlation period (characteristic time) of the covariance model.

Term $\sigma^*[t_k | \mathbf{x}^{(s)}]$ of Eq. (6) denotes an estimate of the conditional standard deviation of the RV $X(t_k)$ given the vector $\mathbf{x}^{(s)}$, that is, $\sigma^*[t_k | \mathbf{x}^{(s)}] \cong \sqrt{\text{Var}\{X(t_k) | \mathbf{x}^{(s)}\}}$. This estimate is written as

$$\sigma^*[t_k | \mathbf{x}^{(s)}] = \left[s_{c_k}^2(t_k) - \sum_{n=1}^{N} \xi_n(k) C_R(t_n - t_k) \right]^{1/2}. \quad (9)$$

Term $w^{(s)}(t_k)$ of Eq. (6) is a simulated realization from a zero-mean, unit-variance, RV $W(t_k)$, which is generated independently from one time step to another:

$$w^{(s)}(t_k) = F_W^{-1}[v^{(s)}(t_k)], \qquad (10)$$

where $v^{(s)}(t_k)$ is a random number uniformly distributed in the interval $[0, 1]$, and $F_W^{-1}(\ )$ denotes the quantile function, that is, the inverse cumulative distribution function (cdf) of $W(t_k)$.

It can be seen that the only random term in Eq. (6) is the simulated quantile $w^{(s)}(t_k) = F_W^{-1}[v^{(s)}(t_k)]$. If $F_W(\ )$

is the standard normal (zero mean, unit variance) Gaussian cdf, then the resulting simulated realization $\{x^{(s)}(t_k), k = 1, \ldots, K\}$ is a realization of a multivariate Gaussian stochastic process. In essence, certain properties of the simulated realizations, for example, their histogram, are controlled by the cdf $F_W(\ )$, or equivalently its quantile function $F_W^{-1}(\ )$. The distributional type of $F_W(\ )$ does not affect the covariance of the simulated realizations, which can be of any type (not only exponential as in first-order autoregressive processes). The only requirement is that $F_W(\ )$ has mean zero and variance one, and that SK [Eqs. (7) through (9)] is used to derive $m^*[t_k | \mathbf{x}^{(s)}]$ and $\sigma^*[t_k | \mathbf{x}^{(s)}]$ (Journel 1999; Caers 2000).

In this paper, the distributional type of $F_W(\ )$ is identified in a nonparametric way to that of forecast intensity values for the month c_k in which day t_k belongs. In other words, the cdf of forecast intensity values is rescaled to a zero-mean, unit-variance, cdf whose quantile function is then used in place of $F_W^{-1}(\ )$ in Eq. (10). In this way, the shape of the histogram of simulated intensity values closely approximates that of the forecasted precipitation. Similar to the binary case of Eq. (2), Eq. (6) can be viewed as an Nth-order autoregressive representation with a random order for visiting the K time steps.

A set of S simulated precipitation intensity realizations $\{x^{(s)}(t_k), k = 1, \ldots, K; \mathbf{b}\}, s = 1, \ldots, S$ can then be generated, where $\mathbf{b}$ denotes an $(L_X \times 1)$ vector of parameter values $\mathbf{b} = [b_1, \ldots, b_{L_X}]$ characterizing the continuous occurrence process $\{X(t), t \in K\}$. For the case of a season-long forecast period, $L_X = 7$, there are seven entries in $\mathbf{b}$: six for the mean and variance of wet-day precipitation intensity within each month, and one last entry for the correlation period η. For simplicity, the correlation period of precipitation intensity is again considered to be constant within the season.

Simulated realizations of precipitation occurrence and intensity are finally combined via Eq. (1) to yield S synthetic precipitation records $\{z^{(s)}(t_k), k = 1, \ldots, K; \mathbf{q}\}, s = 1, \ldots, S$. These S realizations are indexed by a $(L \times 1)$ parameter vector $\mathbf{q} = [\mathbf{a}'\mathbf{b}']'$, with $L = L_I + L_X$, which has as entries the L parameters characterizing the temporal distribution of combined precipitation values $\{z(t_k), k = 1, \ldots, K\}$. Such parameters include the proportion of wet days $p_{c_k}(t_k)$, mean $m_{c_k}(t_k)$, and variance $s^2_{c_k}(t_k)$ of precipitation intensity in any month in which day t_k belongs, as well as the correlation period δ and η of the covariance models $C_{R_I}(\tau; \delta)$ and $C_R(\tau; \eta)$.

It should be noted that we do not impose exact reproduction of monthly aggregated precipitation statistics, for example, p_c, m_c, and s_c. Consequently, we do not impose exact conservation of the total precipitation given by the forecast (or the observed) record. Reproduction of such monthly aggregated statistics is achieved on average, that is, in expected value over a large number of simulated realizations. Exact conservation of total precipitation can be achieved by accounting for this total value in the system of normal

equations in Eq. (8), similar to Valencia and Schaake (1973); variations of this approach are known as disaggregation procedures in stochastic hydrology, and the reader is referred to Salas (1993) for further details.

3. Parameter uncertainty: Bayesian analysis

In the context of propagating uncertainty in dynamically downscaled seasonal forecasts of daily precipitation to hydrologic response studies, the parameters of the stochastic precipitation model can be determined from the forecasted daily precipitation record $\{y(t_k), k = 1, \ldots, K\}$ obtained from a regional climate model (see Fig. 1). In this case, the synthetic precipitation records reproduce important characteristics of the forecast record $\{y(t_k), k = 1, \ldots, K\}$, such as serial correlation and seasonality in both intermittence and precipitation intensity.

By arranging the L forecast-based parameters of the stochastic precipitation model in an $L \times 1$ column vector $\mathbf{q}_f$, the corresponding synthetic precipitation series can be denoted as $\{z^{(s)}(t_k), k = 1, \ldots, K; \mathbf{q}_f\}$, to explicate their dependence on the parameter vector $\mathbf{q}_f$. If the true (unknown) precipitation record $\{z(t_k), k = 1, \ldots, K\}$ for the forecast time period can be adequately represented by the stochastic precipitation model described in this section, then the true record can be regarded as a realization from that model, with a specific (but unknown) parameter vector $\mathbf{q}$:

$$\{z(t_k), k = 1, \ldots, K\} \cong \{z^{(s)}(t_k), k = 1, \ldots, K; \mathbf{q}\}.$$

The procedure of using the forecast-based parameter vector $\mathbf{q}_f$ to generate synthetic realizations of daily precipitation via the stochastic model, entails

$$\{z(t_k), k = 1, \ldots, K\} \cong \{z^{(s)}(t_k), k = 1, \ldots, K; \mathbf{q}_f\}$$

or, equivalently, $\mathbf{q} \cong \mathbf{q}_f$.

The forecast parameter vector $\mathbf{q}_f$, however, is itself uncertain given that it is derived from a set of uncertain forecast values $\{y(t_k), k = 1, \ldots, K\}$. The objective is then to provide a model of uncertainty regarding the true parameter vector $\mathbf{q}$ and to propagate such uncertainty to hydrologic response variables, for example, streamflow calculated via a deterministic model.

We adopt the Bayesian framework (Box and Tiao 1973; Gelman et al. 1995) for characterizing uncertainty in the parameters of the stochastic precipitation model, that is, for providing a model of uncertainty regarding the true (unknown) parameter vector $\mathbf{q}$. Bayesian analysis allows for modeling this parameter uncertainty and evaluating its consequences on related hydrologic response calculations. The true parameter vector $\mathbf{q}$ is regarded as a random vector with a specified *joint* probability distribution. In other words, the L elements of $\mathbf{q}$ are viewed as a *joint* outcome of a collection of L RVs $\mathbf{Q} = [Q_l, l = 1, \ldots, L]'$ modeling uncertainty in the L parameters of the stochastic precipitation model. A *prior* (climatological) joint probability distribution is
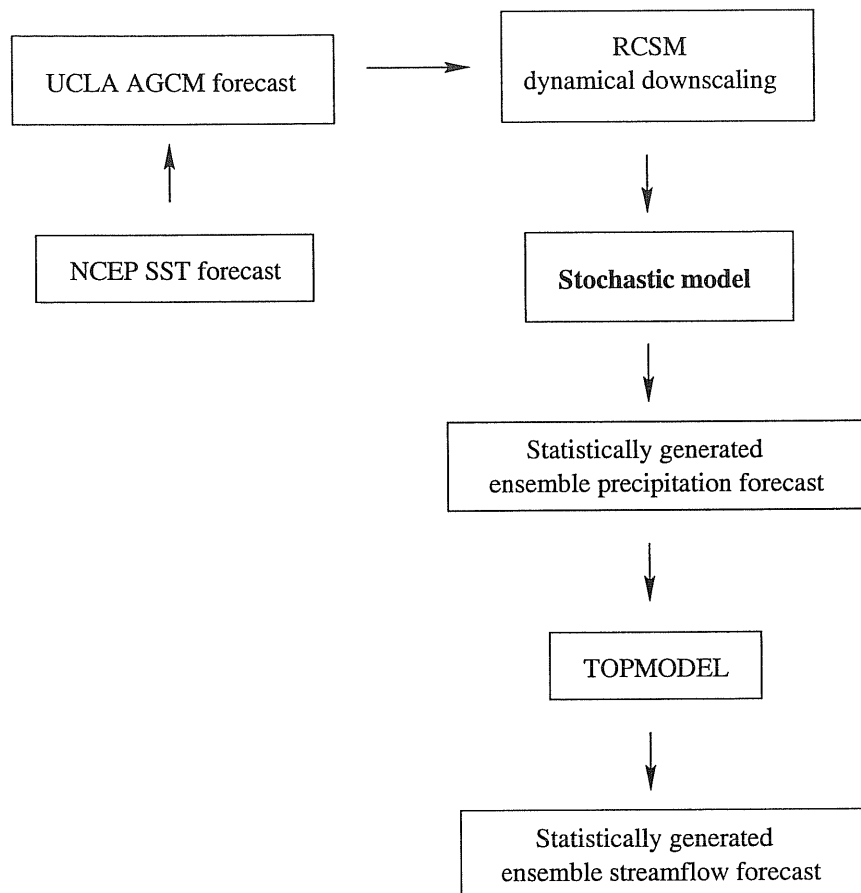
FIG. 1. Schematic representation of climate state forecasts used in the uncertainty propagation procedure.

first established from a set of corresponding parameter vectors derived from historical precipitation data. Bayes's relation is then used for updating (revising) the prior distribution to account for new information; prior knowledge is thus modified in light of the forecast parameter vector $\mathbf{q}_f$.

In the context of precipitation forecasting, prior information regarding the random vector $\mathbf{q}$ can be derived from historical precipitation records at the particular study basin. Consider the case where a historical precipitation record of length N years is available. A set of N (L dimensional) historical parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$, one for each year n, can be then constructed. Each parameter vector $\mathbf{q}_h(n)$ can be regarded as a realization from a common parameter distribution for the N years.

The decision of what constitutes relevant historical information for arriving at prior parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$ should be guided by scientific knowledge. For example, if it is postulated that the forecast period is affected by a strong El Niño–Southern Oscillation (ENSO) event, then one could consider as prior information only historical data in those N' years

(out of the total of N) for which the ENSO signal was particularly strong. Alternatively, if the precipitation forecast constitutes a perturbed climate projection, for example, under doubled carbon dioxide conditions, then the prior parameter distribution could be modified to account for possible time trends. These latter modifications of the prior distribution, however, would make the forecast and the prior distribution dependent; in this case more complex Bayesian models are required.

Let $f(\mathbf{q}; \boldsymbol{\theta})$ denote the multivariate (L dimensional) joint probability density function (pdf) modeling uncertainty regarding the parameter vector $\mathbf{q}$:

$$f(\mathbf{q}; \boldsymbol{\theta}) = \text{Prob}\{Q_1 \in q_1 \pm dq_1, \ldots, Q_L \in q_L \pm dq_L; \boldsymbol{\theta}\},$$

(11)

fully specified by a *hyperparameter* vector $\boldsymbol{\theta}$. The term hyperparameter implies that $\boldsymbol{\theta}$ characterizes the uncertainty in the stochastic precipitation model parameter vector $\mathbf{q}$.

The exact value of this hyperparameter vector $\boldsymbol{\theta}$ is unknown; instead, it is estimated from N historical parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$ and the forecast

parameter vector $\mathbf{q}_f$. Similar to the case of the true parameter vector $\mathbf{q}$, the hyperparameter vector $\boldsymbol{\theta}$ is treated as random.

Let $f(\boldsymbol{\theta})$ denote the prior density function modeling uncertainty regarding the true hyperparameter vector $\boldsymbol{\theta}$. The term prior implies that the hyperparameters of this density are derived from historical parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$ prior to including the information brought by the forecast vector $\mathbf{q}_f$. Bayes's relation allows updating this prior uncertainty model $f(\boldsymbol{\theta})$ to a *posterior* one, when a forecast parameter vector $\mathbf{q}_f$ becomes available:

$$f(\boldsymbol{\theta} | \mathbf{q}_f) = \frac{f(\mathbf{q}_f | \boldsymbol{\theta}) f(\boldsymbol{\theta})}{f(\mathbf{q}_f)}, \qquad (12)$$

where $f(\boldsymbol{\theta} | \mathbf{q}_f)$ denotes the posterior pdf of the hyperparameters $\boldsymbol{\theta}$, that is, their joint pdf after the forecast parameter vector $\mathbf{q}_f$ has been taken into account. The density $f(\mathbf{q}_f | \boldsymbol{\theta})$ is the pdf of the forecast parameter vector $\mathbf{q}_f$ given the hyperparameters $\boldsymbol{\theta}$, and it is termed a likelihood function in Bayesian analysis. The density $f(\mathbf{q}_f)$ is the marginal density of $\mathbf{q}_f$, evaluated as $f(\mathbf{q}_f) = \int f(\mathbf{q}_f | \boldsymbol{\theta}) f(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$.

In this paper, we adopt a specific form of the density functions involved in Eq. (12), which allows for a convenient analytical derivation of the posterior density $f(\boldsymbol{\theta} | \mathbf{q}_f)$. We then examine two situations: (a) the case of an ensemble forecast setting and (b) the case of a single forecast record.

a. The ensemble forecast case

Let $\{y^{(s')}(t_k), k = 1, \ldots, K\}, s' = 1, \ldots, S'$, denote the S' members of an ensemble daily precipitation forecast (Barnett 1995), which are derived from dynamical downscaling based on different large-scale forcing (initial and time-varying conditions) or even on alternative model formulations. We use the superscript s' to distinguish members of ensemble forecasts based on dynamical downscaling from those based on stochastic simulation. From this set of S' alternative ensemble forecast records $\{y^{(s')}(t_k), k = 1, \ldots, K\}, s' = 1, \ldots, S'$, one can construct a corresponding set of S' alternative forecast parameter vectors $\{\mathbf{q}_f(s'), s' = 1, \ldots, S'\}$.

In what follows, we assume that the alternative forecast parameter vectors $\{\mathbf{q}_f(s'), s' = 1, \ldots, S'\}$ are realizations from an L-variate normal (Gaussian) pdf; see, for example, Mardia et al. (1979):

$$f[\mathbf{q}_f(1), \ldots, \mathbf{q}_f(S')]$$

$$= g(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$= (2\pi)^{-L/2} |\boldsymbol{\Sigma}|^{-S'/2}$$

$$\times \exp\left\{ -\frac{1}{2} \sum_{s'=1}^{S'} [\mathbf{q}_f(s') - \boldsymbol{\mu}]' \boldsymbol{\Sigma}^{-1} [\mathbf{q}_f(s') - \boldsymbol{\mu}] \right\},$$

$$(13)$$

where $g(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multi-Gaussian pdf with hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\Sigma}^{-1}$ denote the determinant and inverse of matrix $\boldsymbol{\Sigma}$, respectively.

The L-variate normal pdf [Eq. (13)] is fully specified by two hyperparameters, that is, the $L \times 1$ mean vector $\boldsymbol{\mu}$,

$$\boldsymbol{\mu} = E\{\mathbf{Q}\}, \qquad (14)$$

and its $L \times L$ covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = \text{Cov}\{Q_l, Q_{l'}\} = E\{(\mathbf{Q} - \boldsymbol{\mu})(\mathbf{Q} - \boldsymbol{\mu})'\}, \qquad (15)$$

where the diagonal entries of $\boldsymbol{\Sigma}$ quantify the variance $\text{Var}\{Q_l\}$ of the lth parameter Q_l, and the off-diagonal entries of $\boldsymbol{\Sigma}$ quantify the linear correlation between parameters q_l and $q_{l'}$. In an ensemble forecast setting, the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be regarded as the ensemble average and covariance values of a large number of S' alternative forecast parameter sets $\{\mathbf{q}_f(s'), s' = 1, \ldots, S'\}$.

Because the true values of the above hyperparameters are unknown, the goal of Bayesian analysis, as expressed in Eq. (12), is to provide a model of uncertainty regarding the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the multi-Gaussian density [Eq. (13)], based on prior (historical) parameters $\mathbf{q}_h$ and the forecast parameter vector $\mathbf{q}_f$:

$$f[\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{q}_f(1), \ldots, \mathbf{q}_f(S')]$$

$$= \frac{f[\mathbf{q}_f(1), \ldots, \mathbf{q}_f(S') | \boldsymbol{\mu}, \boldsymbol{\Sigma}] f(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{f[\mathbf{q}_f(1), \ldots, \mathbf{q}_f(S')]}, \qquad (16)$$

where $f[\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{q}_f(1), \ldots, \mathbf{q}_f(S')]$ denotes the posterior density of the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given the S' forecast vectors $\{\mathbf{q}_f(s'), s' = 1, \ldots, S'\}$; $f[\mathbf{q}_f(1), \ldots, \mathbf{q}_f(S') | \boldsymbol{\mu}, \boldsymbol{\Sigma}]$ denotes the likelihood function of the forecast vectors; $f(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the prior density of the hyperparameters; and $f[\mathbf{q}_f(1), \ldots, \mathbf{q}_f(S')]$ denotes the joint density of the forecast vectors.

Consider the case whereby the analyst has full confidence in a measure of variability among the members of the ensemble forecast vectors $\{\mathbf{q}_f(s'), s' = 1, \ldots, S'\}$ and suspects a possible bias of unknown magnitude, common in all the members of the ensemble. In this case, the covariance matrix $\boldsymbol{\Sigma}$ is fixed (known) and the mean vector $\boldsymbol{\mu}$ is random (unknown), entailing that the posterior density [Eq. (16)] becomes $f[\boldsymbol{\mu} | \mathbf{q}_f(1), \ldots, \mathbf{q}_f(S'), \boldsymbol{\Sigma}]$.

A suitable prior density for the single random hyperparameter $\boldsymbol{\mu}$ is the multivariate normal density

$$f(\boldsymbol{\mu}) = g(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h), \qquad (17)$$

where $\boldsymbol{\mu}_h$ is the prior (climatological) mean vector

$$\boldsymbol{\mu}_h = E\{\mathbf{Q}_h\}, \qquad (18)$$

and $\boldsymbol{\Sigma}_h$ is the prior (climatological) covariance matrix

$$\boldsymbol{\Sigma}_h = E\{(\mathbf{Q}_h - \boldsymbol{\mu}_h)(\mathbf{Q}_h - \boldsymbol{\mu}_h)'\}, \qquad (19)$$

both estimated from the historical parameter vectors $\mathbf{q}_h$.

If a multivariate normal density is not appropriate for

the joint density of the historical parameters, a first approximation can be obtained by transforming the N historical values $\{q_h^l(n), n = 1, \ldots, N\}$ of each parameter q_h^l so that they follow a normal density with the same mean and variance. The one-to-one, invertible, transformation used is termed normal scores transform (Deutsch and Journel 1998). A multivariate normal distribution can then be postulated for the resulting normal score transformed values. Simulated parameter realizations can be drawn from this multivariate normal distribution with specified mean and covariance matrix; the latter is calculated from the normal score transformed data. The simulated parameter realizations can be finally back-transformed to approximate the original (non-Gaussian) distribution of historical parameters; the transformation used is the inverse normal scores transform. In this way, the histogram of historical parameter values for any lth parameter is reproduced by construction, as well as the correlation coefficient between values of any two historical parameters q_h^l and $q_h^{l'}$ (in the Gaussian space).

Bayesian techniques also exist for explicit treatment of non-Gaussian distributions, such as those followed by proportions and variances (Gelman et al. 1995). Recall, however, that our goal is to model the posterior multivariate distribution, that is, the *joint* distribution of the L parameters, not a single univariate distribution. Nonparametric kernel density estimation methods and other statistical techniques can be used for modeling such posterior distributions, but they require heavy inference effort and various approximations (Tanner 1996). On the other hand, the multivariate Gaussian distribution is also a (first) approximation, but it calls for a low number of parameters and can be modeled straightforwardly. In addition, a multivariate normal prior density for $\boldsymbol{\mu}$ [Eq. (17)] allows deriving analytically the posterior density for $\boldsymbol{\mu}$ (see below). In this multivariate normal case, the prior density is conjugate with respect to the posterior density, in that they both have the same parametric form (multi-Gaussian in this case). It should be noted, however, that a univariate transformation, such as the normal scores transform, does not entail that the joint distribution of the L sets of transformed variables is multivariate normal. Additional diagnostic checks can be conducted for investigating the appropriateness of the bivariate Gaussian hypothesis, for example, for the normal score variables (Deutsch and Journel 1998).

It can be shown (e.g., Box and Tiao 1973) that the posterior density of the mean vector $\boldsymbol{\mu}$ given a set of S' forecast parameter vectors $\{\mathbf{q}_f(s'), s' = 1, \ldots, S'\}$ (and the fixed covariance matrix $\boldsymbol{\Sigma}_f$) is multivariate normal,

$$f[\boldsymbol{\mu} \mid \mathbf{q}_f(1), \ldots, \mathbf{q}_f(S'), \boldsymbol{\Sigma}_f] = g(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p), \quad (20)$$

with posterior mean $\boldsymbol{\mu}_p$,

$$\boldsymbol{\mu}_p = (\boldsymbol{\Sigma}_h^{-1} + S'\boldsymbol{\Sigma}_f^{-1})^{-1}(\boldsymbol{\Sigma}_h^{-1}\boldsymbol{\mu}_h + S'\boldsymbol{\Sigma}_f^{-1}\overline{\mathbf{q}}_f), \quad (21)$$

where $\overline{\mathbf{q}}_f$ denotes the average forecast parameter vector, and posterior inverse covariance matrix $\boldsymbol{\Sigma}_p^{-1}$:

$$\boldsymbol{\Sigma}_p^{-1} = \boldsymbol{\Sigma}_h^{-1} + S'\boldsymbol{\Sigma}_f^{-1}. \quad (22)$$

The posterior precision matrix $\boldsymbol{\Sigma}_p^{-1}$, that is, the inverse of the posterior covariance matrix $\boldsymbol{\Sigma}_p$, is the sum of the prior precision matrix $\boldsymbol{\Sigma}_h^{-1}$ and the data precision matrix $S'\boldsymbol{\Sigma}_f^{-1}$. The posterior mean vector $\boldsymbol{\mu}_p$ is the weighted average of the prior mean vector $\boldsymbol{\mu}_h$ and the average forecast parameter vector $\overline{\mathbf{q}}_f$, the weights being determined by the precision matrices $\boldsymbol{\Sigma}_h^{-1}$ and $S'\boldsymbol{\Sigma}_f^{-1}$.

Two limiting cases can be distinguished.

1) The prior (historical) information is infinitely more precise than the forecast data, that is, $\boldsymbol{\Sigma}_h^{-1} \gg \boldsymbol{\Sigma}_f^{-1}$. Then, from Eqs. (22) and (21),

$$\boldsymbol{\Sigma}_p^{-1} = \boldsymbol{\Sigma}_h^{-1}$$

$$\Rightarrow \boldsymbol{\mu}_p = \boldsymbol{\mu}_h$$

$$\Rightarrow f[\boldsymbol{\mu} \mid \mathbf{q}_f(1), \ldots, \mathbf{q}_f(S'), \boldsymbol{\Sigma}_f] = g(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h),$$

entailing that the posterior density equals the prior density derived from historical data.

2) The forecast data dominate the prior information, that is, $\boldsymbol{\Sigma}_f^{-1} \gg \boldsymbol{\Sigma}_h^{-1}$, or equivalently $\boldsymbol{\Sigma}_h^{-1} \cong 0$; this is the case of diffuse (completely uncertain) prior information. Then, from Eqs. (22) and (21),

$$\boldsymbol{\Sigma}_p^{-1} = \boldsymbol{\Sigma}_f^{-1}$$

$$\Rightarrow \boldsymbol{\mu}_p = \overline{\mathbf{q}}_f$$

$$\Rightarrow f[\boldsymbol{\mu} \mid \mathbf{q}_f(1), \ldots, \mathbf{q}_f(S'), \boldsymbol{\Sigma}_f] = g(\overline{\mathbf{q}}_f, \boldsymbol{\Sigma}_f),$$

entailing that the posterior density has mean equal to average forecast parameter vector $\overline{\mathbf{q}}_f$ itself.

A special case of this second scenario is that of a quasi-zero forecast covariance matrix, that is, $\boldsymbol{\Sigma}_f \cong \mathbf{0}$. In this situation, the posterior density collapses into a set of L spikes at the L entries of the forecast vector $\mathbf{q}_f$, which corresponds to an infinitely precise (certain) forecast. In an ensemble forecast setting, this would amount to an ensemble of nearly identical forecasts $\{y^{(s')}(t_k), k = 1, \ldots, K\}, s' = 1, \ldots, S'$, that is, a set of nearly identical forecast parameter vectors $\{\mathbf{q}_f(s'), s' = 1, \ldots, S'\}$.

The above results call for knowledge of the forecast covariance matrix $\boldsymbol{\Sigma}_f$. In an ensemble forecast setting, even if such a covariance matrix can be estimated from the various members of the ensemble, it could be still regarded as unknown, in that its exact value cannot be determined with certainty. Worse, in the case of a single member forecast, the forecast covariance matrix $\boldsymbol{\Sigma}_f$ is unobservable, that is, it cannot be estimated from multiple members of the ensemble. Bayesian analysis, however, can be applied in the multivariate normal case with both mean vector $\boldsymbol{\mu}$ *and* covariance matrix $\boldsymbol{\Sigma}$ unknown.

b. The case of a single forecast record

Consider the case of a single forecast precipitation record $\{y(t_k), k = 1, \ldots, K\}$, that is, a single forecast parameter vector $\mathbf{q}_f$. The historical precipitation records $\{z^{(n)}(t_k), k = 1, \ldots, K\}$, $n = 1, \ldots, N$ at the study basin can still be used to derived a set of N historical parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$. The only difference with the previous case of ensemble forecasting (section 3a) is that both the mean vector $\boldsymbol{\mu}$ *and* the covariance matrix $\boldsymbol{\Sigma}$ are unknown. In a multivariate normal setting, one has to assign a suitable joint prior density $f(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ based on the (possibly normal score transformed) historical parameters.

If a multivariate Gaussian density is postulated for these (transformed) historical parameters, then their covariance matrix $\boldsymbol{\Sigma}_h$ follows a Wishart density, that is, a multivariate generalization of the scaled χ^2 density (Mardia et al. 1979). Consequently, we can adopt the inverse Wishart density for modeling the prior density of the $L \times L$ covariance matrix $\boldsymbol{\Sigma}$:

$$f(\boldsymbol{\Sigma}) = \text{Wishart}_{\nu_h}^{-1}(\boldsymbol{\Sigma}_h^{-1})$$

$$= \left[2^{\nu_h L/2}\, \pi^{L(L-1)/4} \prod_{l=1}^{L} \Gamma\!\left(\frac{\nu_h + 1 - l}{2}\right)\right]^{-1}$$

$$\times\ |\boldsymbol{\Sigma}_h|^{\nu_h/2}|\boldsymbol{\Sigma}|^{-(\nu_h + L + 1)/2}\, \exp\!\left[-\frac{1}{2}\, \text{tr}(\boldsymbol{\Sigma}_h \boldsymbol{\Sigma}^{-1})\right], \tag{23}$$

where $\text{Wishart}_{\nu_h}^{-1}(\ )$ denotes the inverse Wishart density with $\nu_h \geq L$ the degrees of freedom, $\text{tr}(\ )$ denotes the trace of a matrix, and $\Gamma(\ )$ denotes the gamma function (Abramovitz and Stegun 1972).

Conditional on outcomes of $\boldsymbol{\Sigma}$ from this inverse Wishart density [Eq. (23)], the density of $\boldsymbol{\mu}$ is multivariate normal:

$$f(\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}) = g(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}/N). \tag{24}$$

Equations (23) and (24) correspond to the joint prior density for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-[(\nu_h + L)/2 + 1]} \exp\!\left[-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}_h \boldsymbol{\Sigma}^{-1})\right.$$
$$\left. -\frac{N}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_h)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_h)\right],$$

with the values ν_h, $\boldsymbol{\Sigma}_h$, N, and $\boldsymbol{\mu}_h$ being inferred from historical parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$.

The posterior density $f(\boldsymbol{\mu}, \boldsymbol{\Sigma} \,|\, \mathbf{q}_f)$, conditional on the forecast parameter vector $\mathbf{q}_f$, has the same form as the prior density (e.g., Gelman et al. 1995) with $\nu_h + 1$ degrees of freedom. The posterior mean vector $\boldsymbol{\mu}_p$ is

$$\boldsymbol{\mu}_p = \frac{N}{N+1}\boldsymbol{\mu}_h + \frac{1}{N+1}\mathbf{q}_f, \tag{25}$$

and the posterior covariance matrix $\boldsymbol{\Sigma}_p$ is

$$\boldsymbol{\Sigma}_p = \boldsymbol{\Sigma}_h + \frac{N}{N+1}(\mathbf{q}_f - \boldsymbol{\mu}_h)(\mathbf{q}_f - \boldsymbol{\mu}_h)'. \tag{26}$$

It can be seen from Eqs. (25) and (26) that the posterior hyperparameters in this case are derived by pooling the forecast parameter vector $\mathbf{q}_f$ with the historical parameters $\mathbf{q}_h$ in a common dataset and calculating its multivariate mean $\boldsymbol{\mu}_p$ and covariance matrix $\boldsymbol{\Sigma}_p$.

In summary, under the multivariate normal model with unknown mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_h$, the Bayesian paradigm for characterizing uncertainty in the single forecast parameter vector $\mathbf{q}_f$ proceeds as follows.

1) Establish the prior mean vector $\boldsymbol{\mu}_h$ and the prior covariance matrix $\boldsymbol{\Sigma}_h$ from the historical data using Eqs. (18) and (19).
2) Compute the posterior mean vector $\boldsymbol{\mu}_p$ and the posterior covariance matrix $\boldsymbol{\Sigma}_p$ from Eqs. (25) and (26).
3) Draw a realization from the joint posterior density $f(\boldsymbol{\mu}, \boldsymbol{\Sigma} \,|\, \mathbf{q}_f)$ in two steps:
   (a) draw a realization of the posterior covariance matrix $\boldsymbol{\Sigma}$ from

   $$f(\boldsymbol{\Sigma} \,|\, \mathbf{q}_f = \text{Wishart}_{\nu_h + 1}^{-1}(\boldsymbol{\Sigma}_p^{-1}),$$

   (b) and, conditional on $\boldsymbol{\Sigma}$, draw a realization of the posterior mean vector $\boldsymbol{\mu}$ from

   $$f(\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \mathbf{q}_f) = g[\boldsymbol{\mu}_p, \boldsymbol{\Sigma}/(N+1)].$$

4) Draw a realization of the parameter vector $\mathbf{q}$ from a multivariate normal density with mean and covariance the previously simulated mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$:

   $$\mathbf{q} \sim g(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

   where the symbol $\sim$ implies that vector $\mathbf{q}$ is a realization from $g(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
5) Repeat steps 3–4 S times to generate S simulated parameter vectors $\{\mathbf{q}(s), s = 1, \ldots, S\}$. For details regarding the procedure of drawing realizations from (inverse) Wishart and multivariate normal densities, see Ripley (1987).
6) Back-transform all simulated parameter values to their respective original (non-Gaussian) marginal distributions.

We now proceed with demonstrating the application of the described uncertainty propagation procedure to a real-world dataset.

4. Case study

The objective is to provide a model of uncertainty regarding forecasted daily streamflow at the Hopland basin in the northern California Coast Range for the winter months (DJF) of 1997/98. The area-averaged daily precipitation forecast used in this case study is a seasonal prediction derived from the Regional Climate System Model (RCSM) (see Kim et al. 2000), and it

spans a period from 1 December 1997 to 28 February 1998, leading to $K = 90$ days. This seasonal forecast was derived using dynamical downscaling nested within a global forecast of the University of California, Los Angeles, (UCLA) Atmospheric General Circulation Model (AGCM), which was in turn forced by a forecast of equatorial sea surface temperature (SST) anomaly made by the National Centers for Environmental Prediction (NCEP); see Fig. 1. The reference area-averaged precipitation record for the same period at Hopland was computed as the weighted average of precipitation values recorded at four nearby rain gauges, namely at Willits, Ukiah, Yorkville, and Lake Mendocino (Miller and Kim 1996). The particular weighting scheme adopted is season specific and is used by the California–Nevada River Forecast Center to derive area-averaged precipitation forcing for operational streamflow predictions.

a. Forecasted daily precipitation and streamflow response

The forecasted and reference (unknown in practice) precipitation records are denoted as $\{y(t_k), k = 1, \ldots, K\}$ and $\{z(t_k), k = 1, \ldots, K\}$, respectively, and are shown in Figs. 2a,c. It can be seen that the forecasted record misses the heavy precipitation event during January (days 32–63). Kim et al. (2000) attributed this mismatch to a poor global forecast of the large-scale forcing, which was used as initial and time-varying lateral boundary conditions to run the RCSM (Fig. 1). Clearly, the statistics of the daily forecasted record are different from those of the corresponding reference record, as indicated by their quantile–quantile plot shown in Fig. 2e; a perfect match in distribution would be depicted by a straight line close to the 45° bisector in this plot. Note the severe underestimation of the low-valued quantiles, indicating that there are many more wet days with precipitation intensity less than about 40 mm day^{-1} in the reference precipitation record than in the forecasted one.

Both daily precipitation records are input into TOPMODEL (Beven and Kirkby 1979) for calculating the associated streamflow response. Calibration of TOPMODEL was based on sensitivity analysis to individual model parameters (Beven and Binley 1992; Kim et al. 1998; Miller et al. 1999). Calibration and verification were performed for two 5-yr time periods that include the 1983 El Niño season (Miller and Kim 1996). In essence, all TOPMODEL parameters are held spatially constant, apart from the topographic index, which takes into account the spatially varying gradient and upstream area of each individual grid node (Miller and Kim 1996; Kim et al. 1998). Kim et al. (2000) reported very good reproduction of observed streamflow from a TOPMODEL-derived response based on hindcasted precipitation forcing, in which the RCSM was forced by the NCEP reanalysis. This result indicates that a correct specification of the input precipitation forcing is the most im-

portant factor in reproducing the observed streamflow via TOPMODEL. Consequently, in what follows we will ignore any uncertainty or error in the physics of the streamflow calculation due to intrinsic nonuniqueness of TOPMODEL parameters, as well as any uncertainty due to the spatial variability of precipitation. In other words, TOPMODEL and its calibrated parameters are viewed as a fixed transfer function for evaluating streamflow response to area-averaged precipitation forcing.

The spatial variability of precipitation can be significant when dealing with large basins with strong elevation gradients. The area of Hopland basin is 658 km^2, and although precipitation rates vary spatially during a given storm event, we are modeling this (low elevation) small-to-medium-sized basin with uniform spatial precipitation as a first approximation. We are focusing on effective properties throughout this study in a semidistributed framework. In addition, spatially uniform precipitation is in agreement with the National Weather Service mean area precipitation, which was used for calibrating TOPMODEL. An ongoing study takes into account the spatial variability of precipitation, as well as hydrologic-scale sensitivity to response times.

The TOPMODEL-derived streamflow response, corresponding to the forecasted and reference precipitation records, and the observed streamflow at Hopland for DJF 1997/98, are shown in Figs. 2b,d. Note the similarity of the observed streamflow to the TOPMODEL-derived response based on the reference precipitation record. This indicates that TOPMODEL-related errors are small when compared with any misspecification of precipitation forcing characteristics. Indeed, this fact is also confirmed by the mismatch between observed and TOPMODEL-derived streamflow, when the latter is based on a misspecified forecasted precipitation record such as the seasonal forecast. This latter misspecified forcing also results in a poor reproduction of the distribution of the observed streamflow record, as indicated by the quantile–quantile plot of Fig. 2f.

Even if the mismatch between forecasted and observed precipitation forcing is small for high values, above 40 mm day^{-1} (Fig. 2e), the corresponding mismatch for high streamflow values is amplified (Fig. 2f). This result is due to the fact that the forecasted precipitation record (Fig. 2a) has different persistence properties from the reference record (Fig. 2c), as well as from nonlinearities in the TOPMODEL formulation. Such a difference in persistence is revealed as a severe underestimation of high flow stage from the forecast-based streamflow. Note that a comparison of observed and forecasted streamflow in terms of distribution characteristics does not inform the performance of the forecast in pinpointing the actual time of high precipitation and corresponding streamflow. Evaluation of such performance is accomplished by examining bivariate statistics, such as the correlation coefficient between the forecasted and observed records. In this case, such cor-
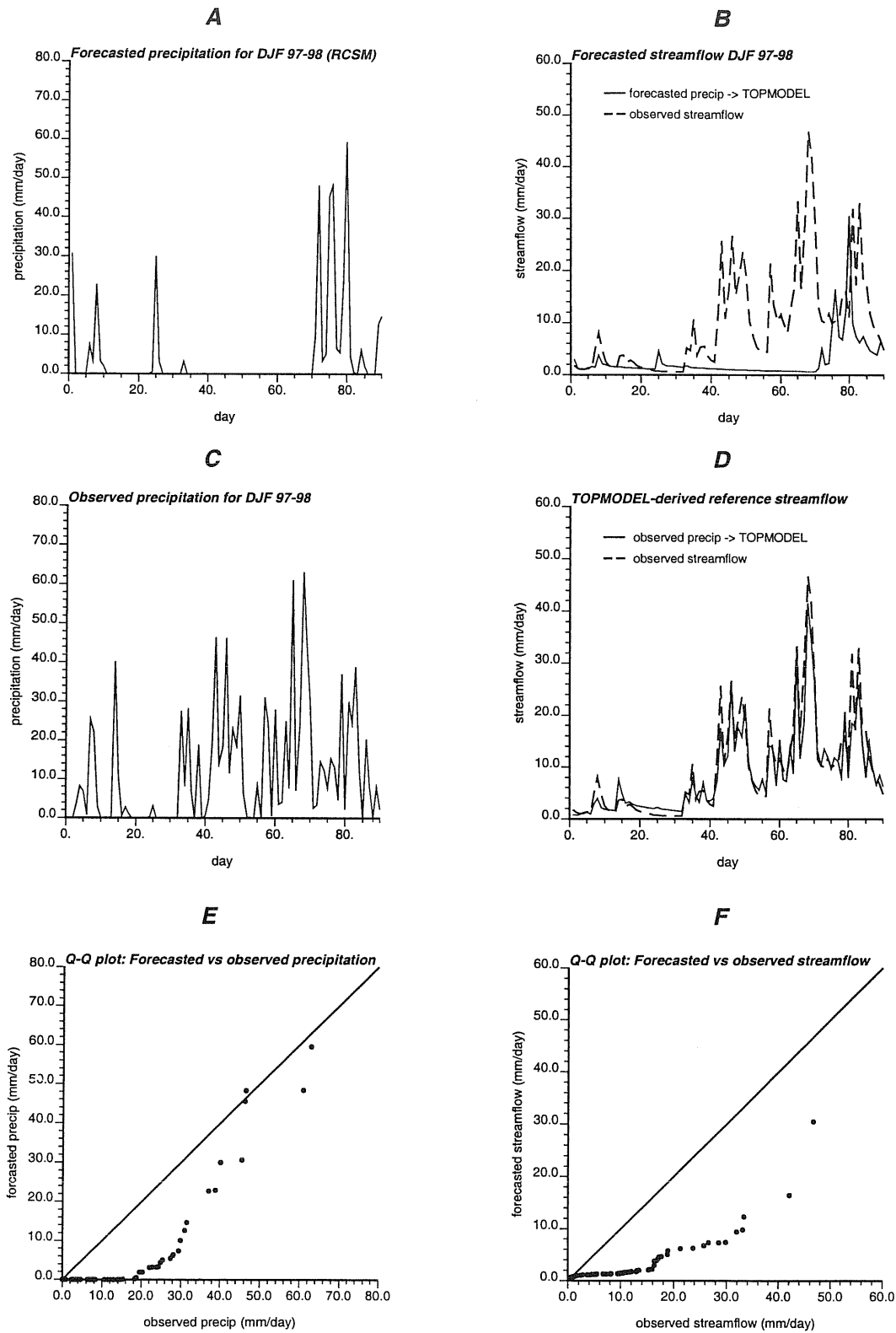
FIG. 2. Streamflow response to precipitation forcing at Hopland basin. (a) Forecasted precipitation from RCSM; (b) corresponding forecast streamflow derived via TOPMODEL (solid line) and observed streamflow record (dashed line); (c) reference precipitation record; (d) corresponding reference streamflow derived via TOPMODEL (solid line) and observed streamflow record (dashed line); (e) and (f) quantile–quantile plots of distributions of forecasted versus reference precipitation, and forecasted versus observed streamflow.

TABLE 1. Parameters of stochastic precipitation model based on forecasted and reference precipitation records: $\{p_D, p_J, p_F\}$ denote the proportions of wet days for DJF 1997/98, $\{m_D, m_J, m_F\}$ denote the corresponding mean wet-day precipitation intensities (mm day^{-1}), $\{s_D, s_J, s_F\}$ denote the corresponding standard deviations of wet-day precipitation intensity (mm day^{-1}), and δ and η denote the seasonal correlation time of the occurrence and intensity anomalies.

Parameter	p_D	p_J	p_F	m_D	m_J	m_F	s_D	s_J	s_F	δ	η
Forecast	0.42	0.10	0.57	7.93	1.10	18.45	11.20	1.50	19.38	4	0
Reference	0.45	0.87	0.93	9.14	15.94	19.95	11.47	13.41	16.99	3	1

relation coefficient values are close to zero for both precipitation and streamflow records.

The mismatch between forecasted and reference precipitation and associated streamflow documented above renders the task of uncertainty assessment regarding forecasted forcing and its propagation to associated streamflow a necessity. In the next section, we postulate that the values of the forecast record themselves are not reliable, yet their averages over appropriate timescales and their serial correlation characteristics constitute useful pieces of information.

b. Stochastic simulation based on forecast parameters

The stochastic precipitation model described in section 2 is employed for generating synthetic precipitation records using parameters derived from the forecast record (Fig. 2a). The set of $L = 11$ forecast-based parameters (vector $\mathbf{q}_f$) for the three winter months (DJF) of 1997/98 is shown in Table 1. These parameters include three proportions of wet days $\{p_D, p_J, p_F\}$, three mean and standard deviation values of wet-day precipitation intensity $\{m_D, m_J, m_F\}$ and $\{s_D, s_J, s_F\}$, and the correlation periods δ and η for the precipitation occurrence and wet-day intensity anomalies, respectively. For the Hopland basin, it was found that the covariance functions for both these processes are well approximated by the exponential family, for example, $C_{R_l}(\tau) = \exp(-3\tau/\delta)$. Similar parameters (vector $\mathbf{q}$) are derived from the reference precipitation record $\{z(t_k), k = 1, \ldots, K\}$ for subsequent comparison (see Table 1). Note the mismatch between the forecasted proportion of wet days in January (0.10) versus that actually observed (0.87). A similar severe mismatch is evident for the mean wet-day precipitation intensity in January (forecast = 1.10 mm day^{-1}; observed = 15.94 mm day^{-1}).

Two (out of $S = 100$) synthetic precipitation records $\{z^{(s)}(t_k), k = 1, \ldots, K; \mathbf{q}_f\}$, generated using forecast-based parameters are shown in Figs. 3a,c. The TOP-MODEL-derived streamflow responses corresponding to the two synthetic precipitation records of Figs. 3a and 3c are shown in Figs. 3b and 3d. Figure 3e gives the 95% streamflow intervals calculated from 100 TOP-MODEL runs on the 100 synthetic precipitation records. The upper and lower probability intervals at any single time step t_k represent the streamflow values that bracket the unknown streamflow response for the *same* time step t_k in 95 out of 100 simulated realizations. Such prob-

ability intervals, and all subsequent ones presented in this study, are therefore pointwise intervals, that is, they pertain to a single time step; they should not be interpreted as joint intervals involving more than one time step. The simulated precipitation and streamflow records (Figs. 3a–d) have similar characteristics with the corresponding forecast records (Figs. 2a,b), and the 95% probability intervals for streamflow bracket the forecast streamflow record (Fig. 3e).

Because the parameters of the stochastic precipitation model are monthly aggregate statistics, one should not expect the simulated realizations to pinpoint (reproduce exactly) the timing of forecasted high or low stage. The time of peak of daily streamflow cannot be captured by a model with monthly aggregated parameters. The only statistics at the daily level are those provided by the correlation times δ and η, which characterize persistence patterns within the entire season and do not provide any information on the actual time of peak of precipitation. This is the reason why the streamflow probability intervals (Fig. 3e) are very similar within each month and differ significantly from one month to another.

The reproduction of input forecast parameter values from the stochastic precipitation model is shown for selected parameters in Fig. 4. The boxplots describe the range of variability of the statistics of the $S = 100$ simulated precipitation records: outside whiskers correspond to the 95% probability intervals, inside boxes to the 50% probability intervals, and the vertical solid line in the box to the median value. It can be seen that the simulated parameter values are centered on the forecast parameter value, indicated with a bullet. Similar reproduction was also obtained for the other parameters (not shown), including the correlation periods δ and η of the covariance models $C_{R_l}(\tau; \delta)$ and $C_R(\tau; \eta)$. As noted in section 2a, reproduction of monthly aggregated statistics, that is, reproduction of the stochastic precipitation model parameters, is achieved on average over a large number of simulated realizations.

Clearly, the forecast parameters, used to generate the synthetic precipitation records for arriving at the streamflow probability intervals of Fig. 3e, are uncertain; even worse, they are different from the parameters of the reference precipitation record. The task is now (a) to provide a model of uncertainty regarding the stochastic precipitation model parameters using parameter values derived from historical daily precipitation records at Hopland and (b) to propagate such parameter uncer-

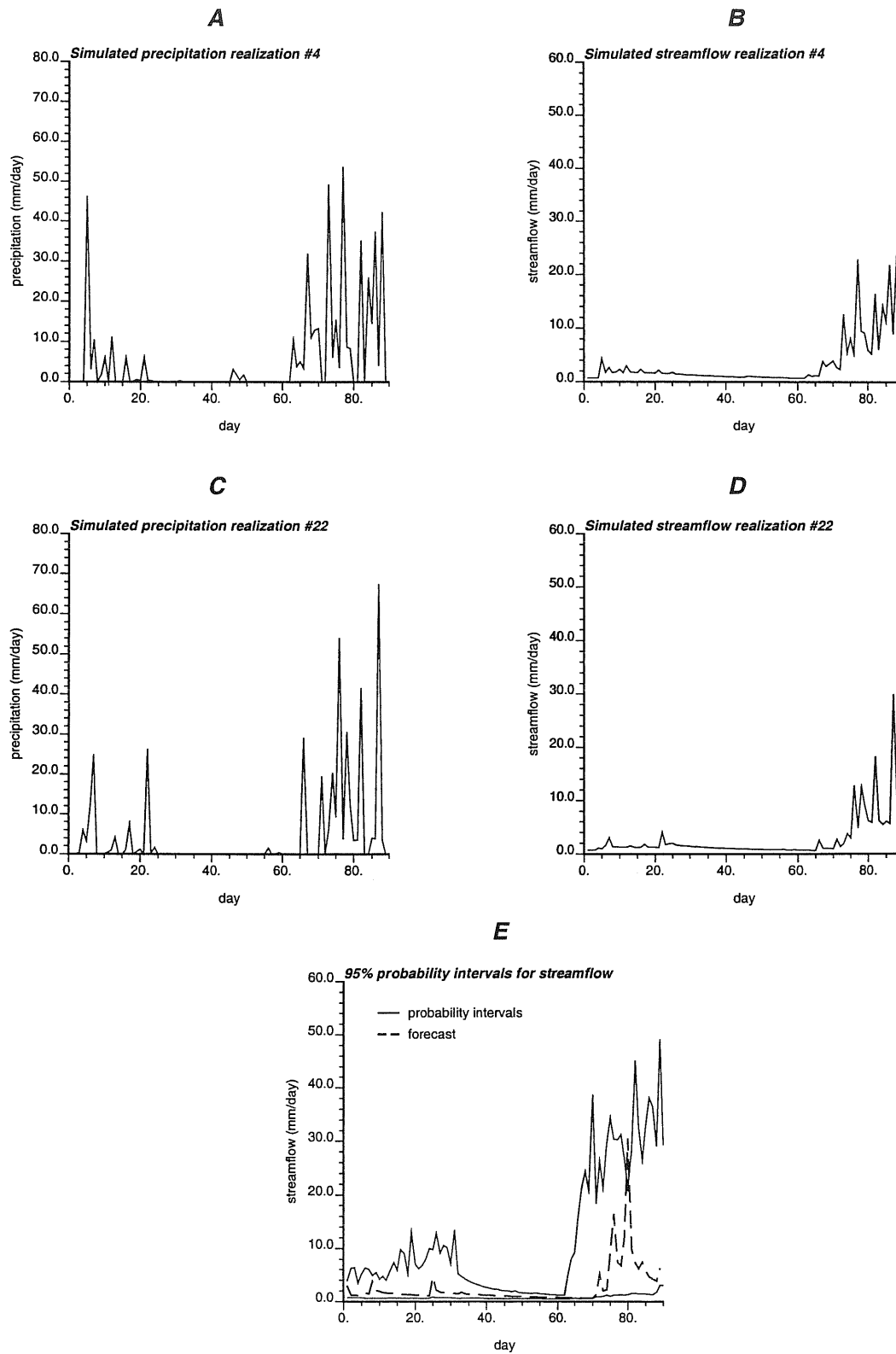FIG. 3. Streamflow response to simulated precipitation forcing based on forecast parameters (diffuse prior information). (a) and (b) Two (out of 100) synthetic precipitation records; (c) and (d) corresponding simulated streamflow realizations derived via TOPMODEL; (e) 95% probability intervals (solid lines) calculated from 100 streamflow realizations, and forecasted streamflow record (dashed line).

A

Simulated wet-day proportion (December)

Number of Data	100
mean	0.42
std. dev.	0.15
coef. of var	0.35
maximum	0.84
upper quartile	0.52
median	0.42
lower quartile	0.32
minimum	0.10

B

Simulated mean intensity (December)

Number of Data	100
mean	8.55
std. dev.	3.18
coef. of var	0.37
maximum	16.75
upper quartile	10.35
median	8.32
lower quartile	6.33
minimum	2.30

C

Simulated wet-day proportion (January)

Number of Data	100
mean	0.10
std. dev.	0.09
coef. of var	0.83
maximum	0.42
upper quartile	0.15
median	0.10
lower quartile	0.03
minimum	0.00

D

Simulated mean intensity (January)

Number of Data	84
number trimmed	16
mean	1.30
std. dev.	0.82
coef. of var	0.63
maximum	3.46
upper quartile	1.98
median	1.09
lower quartile	0.76
minimum	0.02

E

Simulated wet-day proportion (February)

Number of Data	100
mean	0.58
std. dev.	0.15
coef. of var	0.25
maximum	0.89
upper quartile	0.68
median	0.59
lower quartile	0.48
minimum	0.25

F

Simulated mean intensity (February)

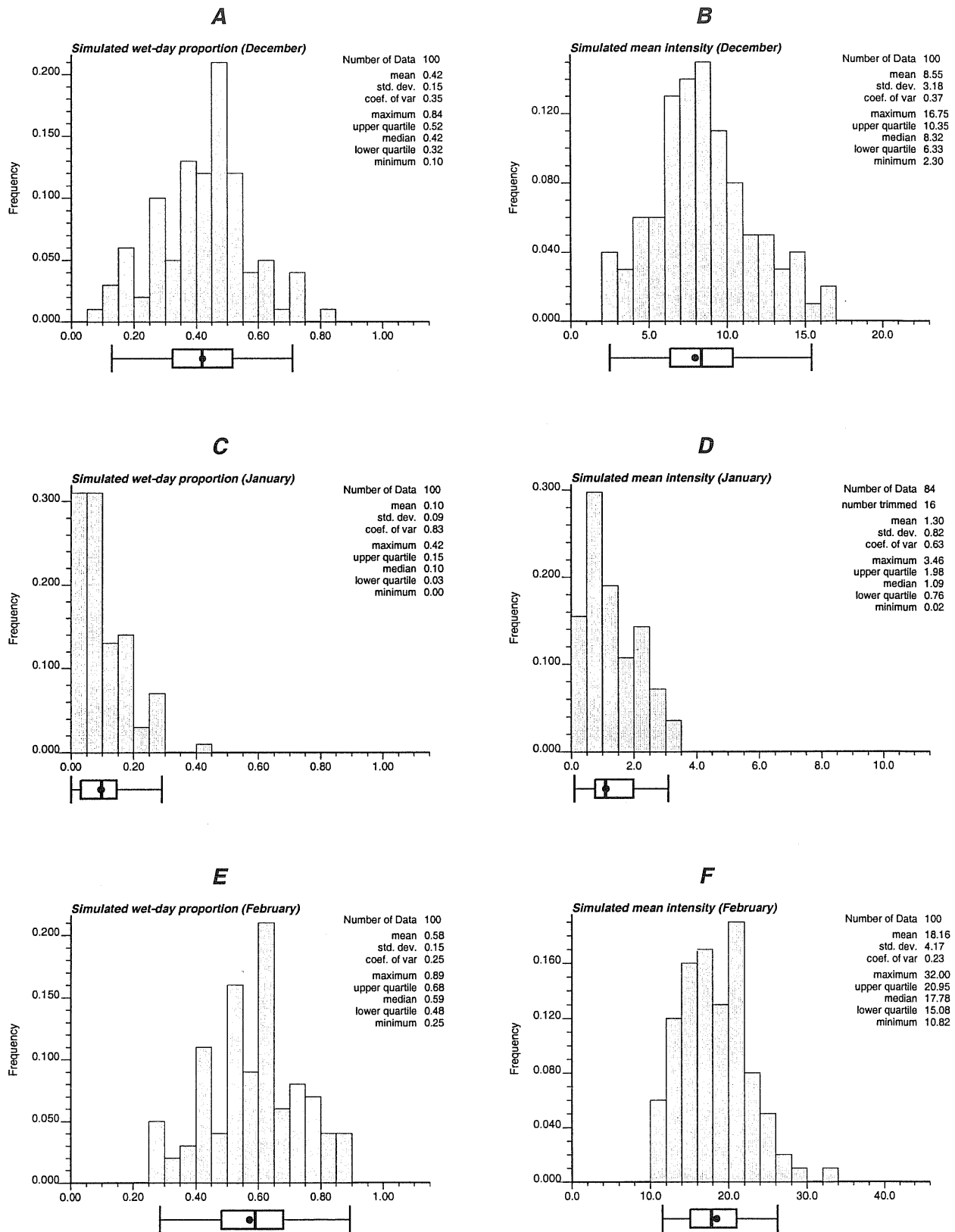|	
Number of Data	100
mean	18.16
std. dev.	4.17
coef. of var	0.23
maximum	32.00
upper quartile	20.95
median	17.78
lower quartile	15.08
minimum	10.82

FIG. 4. Histograms of monthly aggregated statistics from simulated realizations of area-averaged daily precipitation, based on the forecast parameter vector. (a), (c), and (e) Simulated wet-day proportions for DJF; (b), (d), and (f) simulated mean intensities (mm day^{-1}) for DJF. The bullet (●) indicates the corresponding monthly aggregated statistics obtained from the forecast daily precipitation record.

tainty to streamflow response calculations via TOP-MODEL.

c. Stochastic simulation based on historical parameters

We first investigate streamflow uncertainty bounds at Hopland for the case of a diffuse (uninformative) forecast parameter vector $\mathbf{q}_f$ (see section 3a). This scenario corresponds to a forecast with very low skill, whereby the analyst has resorted to the climatological information available at the study basin. In the context of Bayesian analysis, a diffuse forecast implies that the climatological (prior) parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$ constitute the most reliable piece of information.

Historical parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$ for Hopland are derived from daily (area averaged) precipitation records during a period of 34 yr from 1958 to 1992. The histograms of selected parameters are shown in Figs. 5a–f. Again, boxplots describe the range of variability of the statistics of the 34 parameter values: outside whiskers correspond to the 95% probability intervals, inside boxes to the 50% probability intervals, and the vertical solid line in the box to the median parameter value. Symbols ($\bigcirc$) and ($\times$) depict the corresponding parameters derived from the forecasted and reference precipitation records, respectively. From Fig. 5c, it can be seen that the historical wet-day proportions in January do not include the corresponding reference and forecasted proportions. The same is also true for the wet-day proportion in February (Fig. 5e), although in this case the reference parameter lies within the range of variability of the corresponding historical proportions. Clearly, the specific forecast period (DJF 1987/98) was an unusual one in terms of precipitation, in that it was affected by a strong historical ENSO event.

A set of $S = 100$ correlated realizations of $L = 11$ parameters was drawn from the joint density of historical parameters. First, the $N = 34$ historical values $\{q_h^l(n), n = 1, \ldots, N\}$ of each parameter q_h^l were transformed to follow a normal distribution with the same mean and variance as the historical ones [see Eqs. (18) and (19)] via the normal scores transform. Then, a set of $S = 100$ parameter vectors $\{\mathbf{q}_h(s), s = 1, \ldots, S\}$ were drawn from a multivariate normal distribution with mean $\boldsymbol{\mu}_h$ and covariance $\boldsymbol{\Sigma}_h$ (the off-diagonal entries of this covariance matrix were calculated from the normal score transformed data). The simulated parameter realizations were finally back-transformed to approximate the original (non-Gaussian) distribution of historical parameters; the transformation used is the inverse normal scores transform. In this way, the histogram of historical parameters is reproduced by construction, as well as the correlation coefficient between any pair of historical parameters (in the Gaussian space). This histogram reproduction is shown for selected parameters via the quantile–quantile plots of Fig. 6. A similar reproduction was also obtained for the case of the correlation coefficient

between any two pairs of historical parameters (not shown).

These correlated parameter realizations were subsequently used by the stochastic precipitation model (described in section 2) to generate $S = 100$ synthetic precipitation records. Two of these records are shown in Figs. 7a,c, along with the corresponding TOPMODEL-derived streamflow response (Figs. 7b,d). The 95% pointwise probability intervals based on the full set of 100 streamflow realizations are shown in Fig. 7e, along with the observed streamflow record. These latter probability intervals are wider in comparison with those shown in Fig. 3e, because of the introduction of the additional parameter uncertainty.

d. Stochastic simulation based on forecast and historical parameters

We now investigate streamflow uncertainty bounds at Hopland for the case of a nondiffuse (informative) forecast parameter vector $\mathbf{q}_f$ in a single member ensemble forecast setting (see section 3b). In a multivariate normal framework, this implies that both the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are unknown. The additional uncertainty due to the unknown covariance matrix $\boldsymbol{\Sigma}$ is expected to lead to wider uncertainty bounds for the streamflow response at Hopland.

First, the posterior mean vector $\boldsymbol{\mu}_p$ and covariance matrix $\boldsymbol{\Sigma}_p$ are computed from Eqs. (25) and (26), using the N historical parameter vectors $\{\mathbf{q}_h(n), n = 1, \ldots, N\}$ and the single forecast parameter vector $\mathbf{q}_f$. Then, a set of $S = 100$ realizations of parameter vectors $\{\mathbf{q}_h(s), s = 1, \ldots, S\}$ is drawn from the posterior multivariate normal distribution using the procedure described in section 3b.

The simulated parameter vectors were used by the stochastic precipitation model to generate a new set of $S = 100$ synthetic precipitation records, which then were input to TOPMODEL for calculating the corresponding streamflow response. The 95% probability intervals for streamflow are shown in Fig. 8a, along with the observed streamflow record. The probability intervals are much wider in comparison with those shown in Figs. 3e and 7e, because of the introduction of additional parameter uncertainty (unknown covariance matrix). Figure 8b gives the corresponding streamflow probability intervals calculated from 100 TOPMODEL-derived streamflow responses to 100 synthetic precipitation records generated using the reference parameters obtained from the reference precipitation record (see Fig. 2c and Table 1).

Figure 8 indicates that a correct specification of the stochastic precipitation model parameters leads to a more accurate and precise streamflow uncertainty model for Hopland during DJF 1997/98, when compared with the one shown in Fig. 8a. Here, the term accurate implies that the probability intervals bracket the observed streamflow record, and the term precise implies that
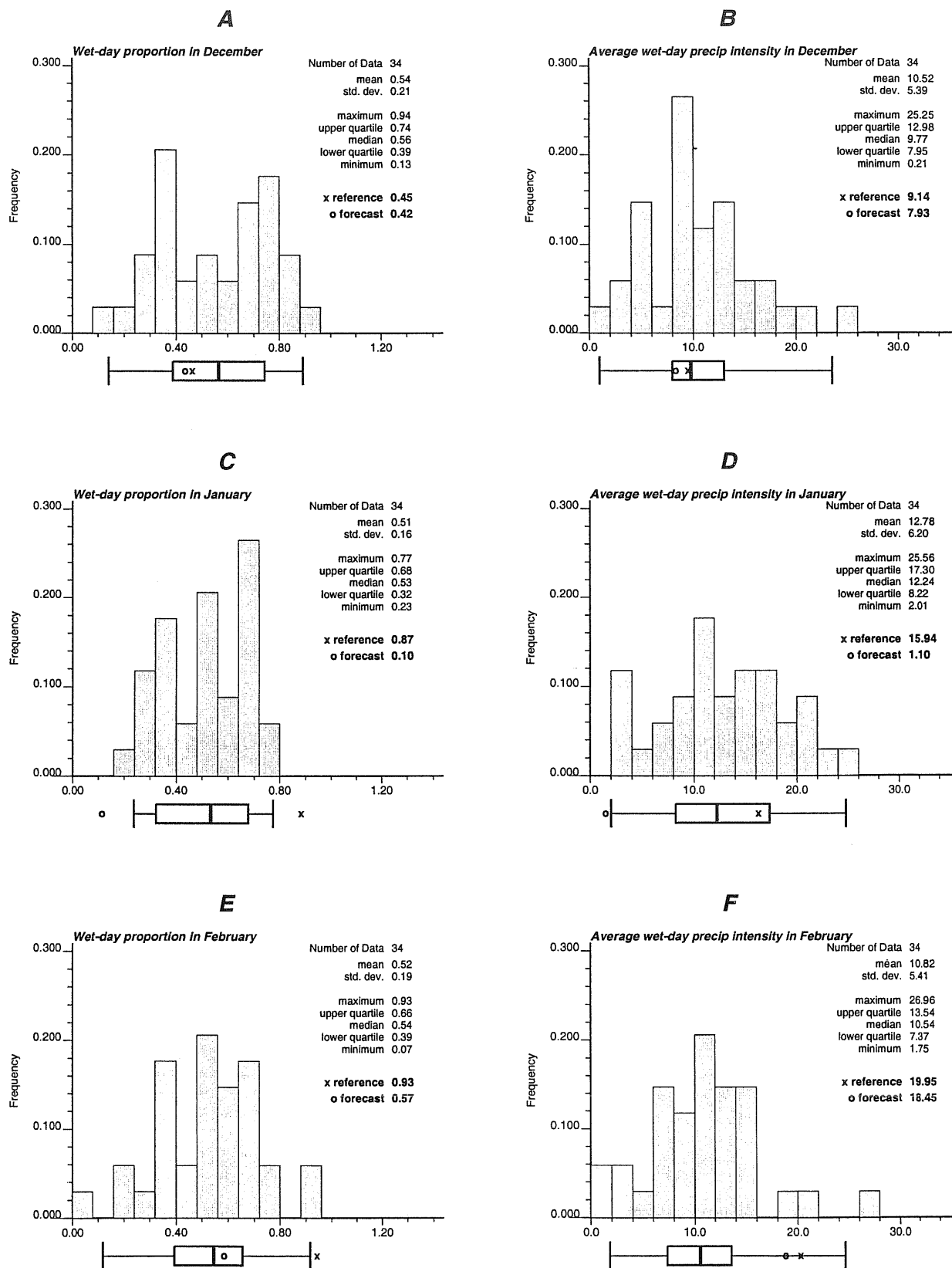
FIG. 5. Histograms of selected stochastic precipitation model parameters derived from historical (1958–92) DJF data at Hopland, along with the corresponding forecast (O) and reference (X) parameters (average wet-day precipitation intensity is expressed in millimeters per day).
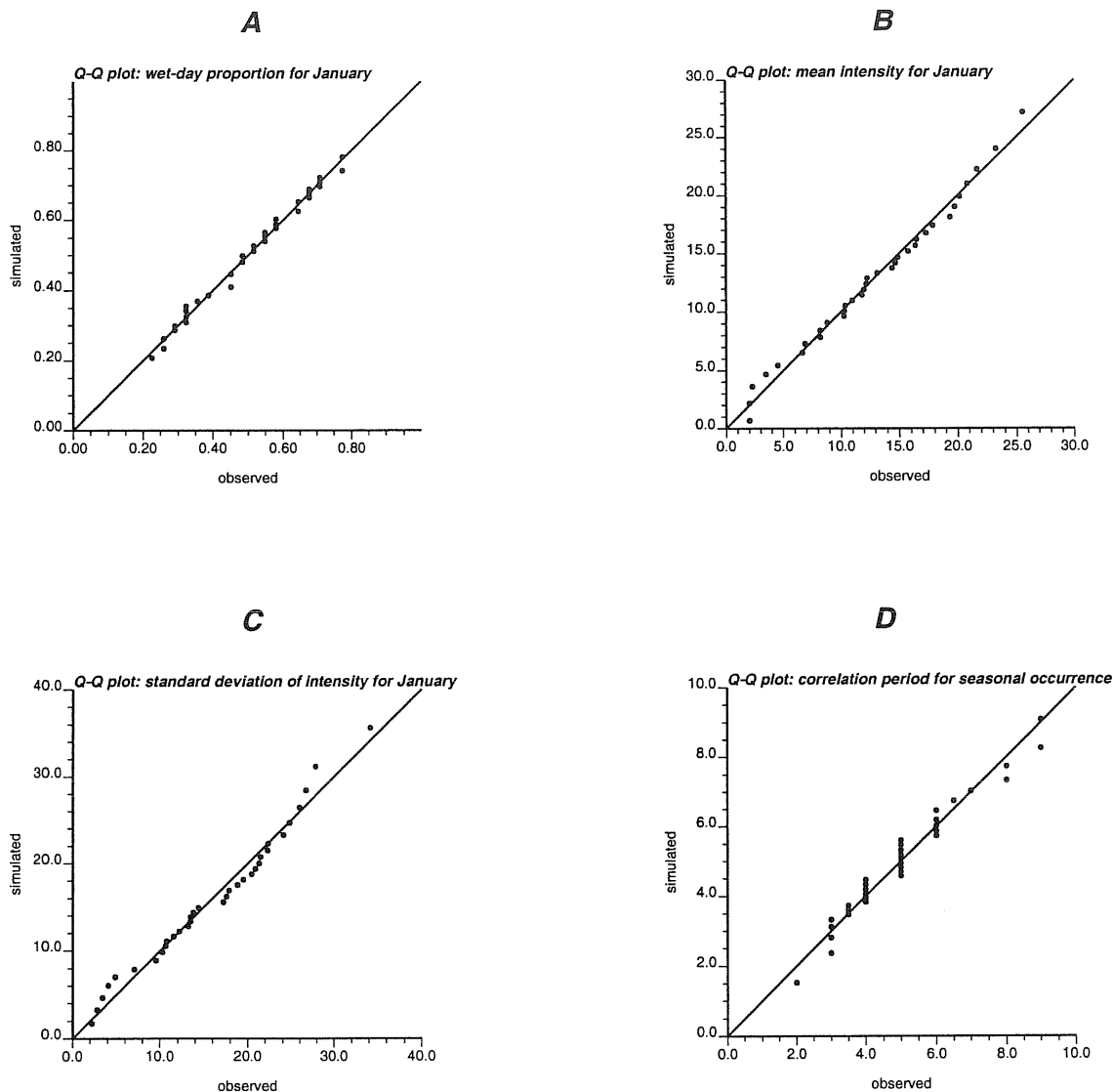
A

Q-Q plot: wet-day proportion for January

B

Q-Q plot: mean intensity for January

C

Q-Q plot: standard deviation of intensity for January

D

Q-Q plot: correlation period for seasonal occurrence

FIG. 6. Reproduction of histograms of historical parameter values from the corresponding simulated realizations. Quantile–quantile plots of distributions of simulated versus historical values for (a) wet-day proportion in Jan, (b) mean intensity in Jan, (c) standard deviation of intensity in Jan, and (d) correlation length of the covariance model $C_{R_l}(\tau; \delta)$.

these intervals are narrow. Simultaneous achievement of accuracy and precision in uncertainty modeling is the ultimate, but unfortunately conflicting, objective: the more confidence one places in a forecast record, the more precise is the uncertainty model and the larger the risk for it to be inaccurate.

e. Uncertainty in forecasted flood probabilities

Last, we investigate uncertainty regarding a flood forecast at Hopland, itself a probability value. For each member of the ensemble sets of simulated streamflow realizations corresponding to different parameter vectors (see sections 4b through 4d), we calculated the percentage of days with simulated stage above the threshold value of 40 mm day^{-1} (also termed crossing rate). The histograms of such crossing rates, which can also be interpreted as flood probabilities, are shown in Figs. 9a–c. We also calculated the corresponding crossing rates from 100 streamflow realizations based on the reference precipitation parameters (see Table 1), and their histogram is shown in Fig. 9d. The crossing rate calculated from the reference TOPMODEL-derived streamflow record of Fig. 2d (solid line) is 1.1% and is depicted in Figs. 9a–d with a bullet; the corresponding forecast-based crossing rate is 0.0%.

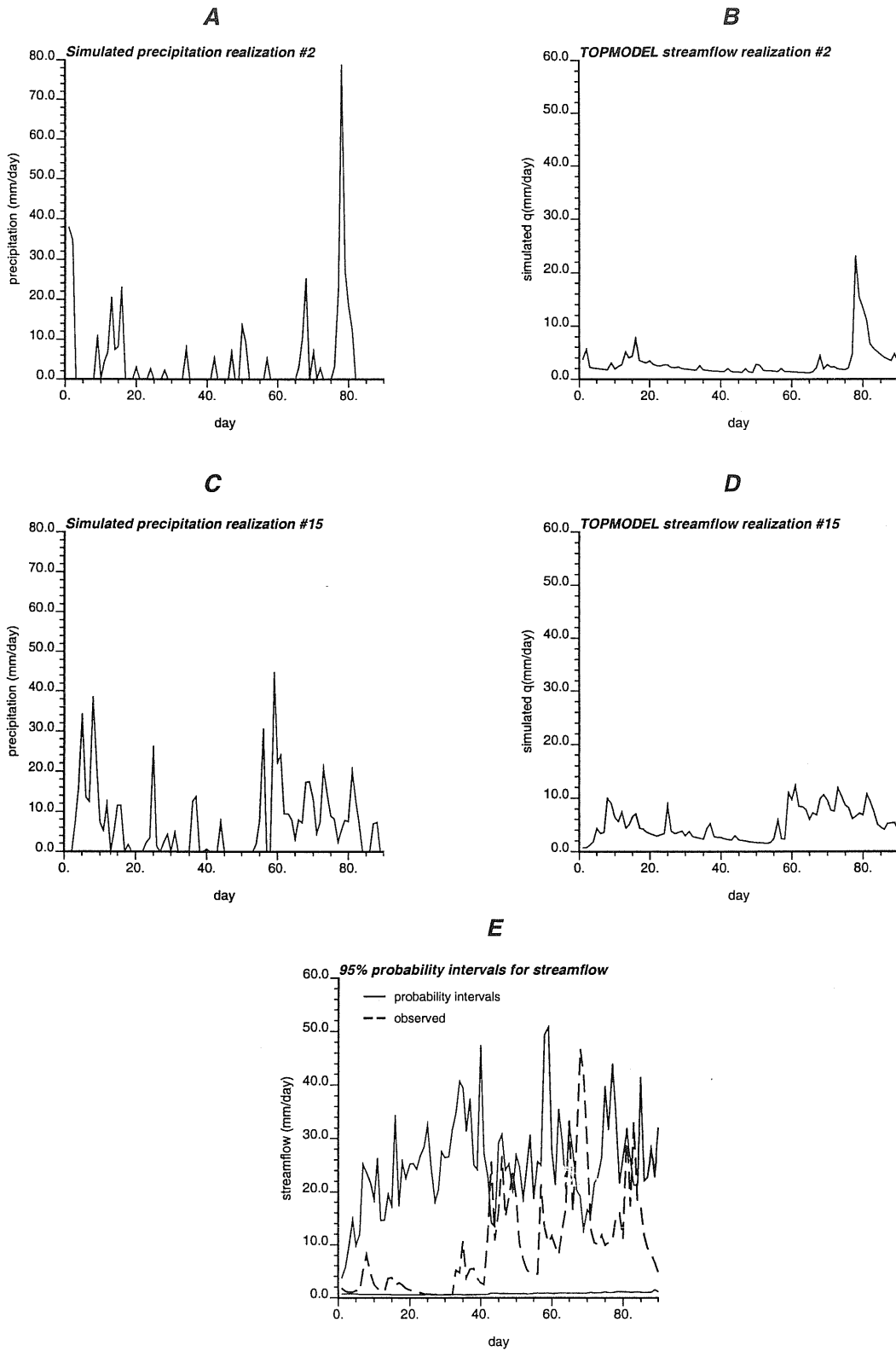Simulation based on forecast parameters (Fig. 9a) leads to a small variability in the simulated rates, with

FIG. 7. Streamflow response to simulated precipitation forcing based on historical parameters (diffuse forecast information). (a) and (c) Two (out of 100) synthetic precipitation records; (b) and (d) corresponding simulated streamflow realizations derived via TOPMODEL; (e) 95% probability intervals (solid lines) calculated from 100 streamflow realizations, and observed streamflow record (dashed line).
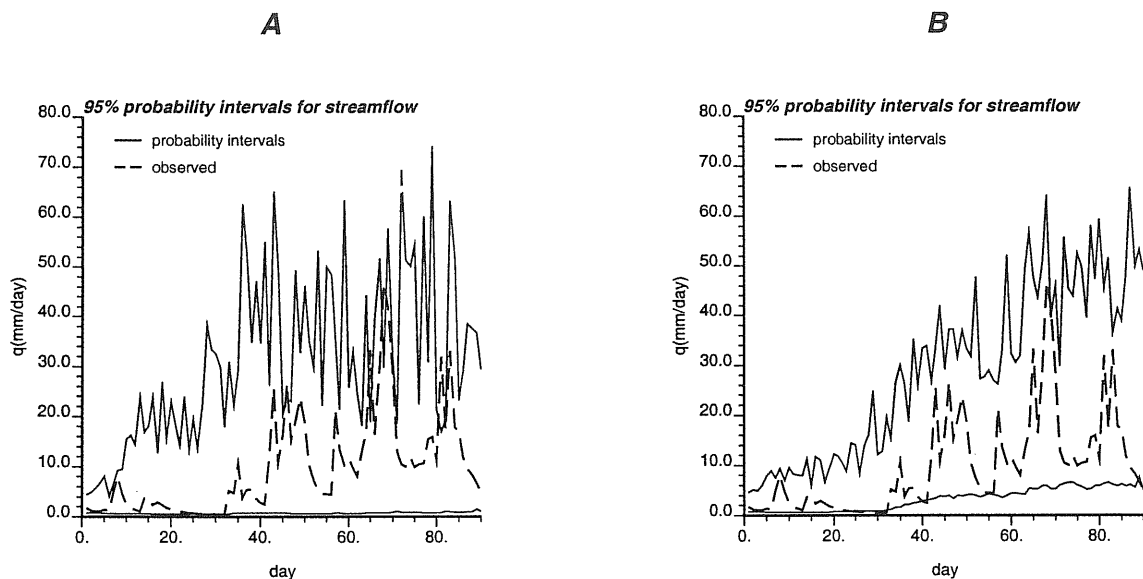
FIG. 8. 95% probability intervals for TOPMODEL-calculated streamflow response to simulated precipitation forcing based on (a) simulated parameters derived from the historical *and* forecast precipitation records and (b) reference parameters derived from the reference precipitation record.

the reference of 1.1% being underestimated (the mean simulated crossing rate is 0.14%). Simulation based on historical parameters (Fig. 9b) leads to an increased variability in the simulated crossing rates, with the reference rate of 1.1% being now less underestimated (the mean simulated crossing rate is 0.51%). The reference rate of 1.1% is still not reproduced, because of the deviation of the reference precipitation model parameters from the range of variability of the corresponding historical parameters. Simulation based on *both* forecast and historical parameters (Fig. 9c) leads to an even larger variability in the simulated crossing rates, with the reference rate of 1.1% being now even less underestimated (the mean simulated crossing rate is now 1.01%). Last, the simulated crossing rates for the case of streamflow simulation based on the reference precipitation model parameters (Fig. 9d) bracket the reference rate of 1.1%, and the mean simulated rate is now 1.12%. This latter exercise indicates that a correct specification of the parameters of the stochastic precipitation model, based on the reference parameter set, leads to relatively more accurate and precise bounds for the crossing rates.

5. Conclusions

A stochastic simulation procedure for propagating uncertainty in dynamically downscaled seasonal forecasts of area-averaged daily precipitation to associated hydrologic response studies is presented in this paper. Synthetic area-averaged daily precipitation records are generated from a stochastic precipitation model whose parameters (e.g., proportion of wet days in a month, mean and variance of wet-day precipitation intensity in a month) are derived from the dynamically downscaled

seasonal forecast. Parameter uncertainty is characterized via an empirical Bayesian model, using as prior information a set of similar climatological parameter values derived from historical precipitation records at the study basin. The impact of uncertain daily precipitation forcing on the associated streamflow response is evaluated by running a deterministic hydrologic model (in this case TOPMODEL) on each member of the ensemble.

The stochastic simulation procedure is applied for assessing uncertainty in forecasted daily streamflow due to uncertain area-averaged daily precipitation forcing at the Hopland basin in the northern California Coast Range for the winter months (DJF) of 1997/98. In the case of correct parameter identification, that is, when stochastic precipitation model parameters are derived from the reference (but unknown in practice) precipitation record at Hopland, it is demonstrated that the developed model leads to realistic uncertainty bounds for the observed streamflow response. When the stochastic precipitation model is driven by the forecast parameter vector only, the uncertainty bounds for the streamflow response bracket the forecasted streamflow response but are too narrow. The risk of incorrect parameter identification, that is, the risk of a poor forecast precipitation record, is mitigated by accounting for historical data at Hopland, leading to much wider probability intervals for streamflow response. Bayesian analysis simply explicates the trade-off between uncertainty bounds from a single, but possible wrong, forecast versus much wider, but more likely to include the true response, uncertainty bounds from historical records.

Although precipitation spatial variability is not accounted for in this study, simulated streamflow closely
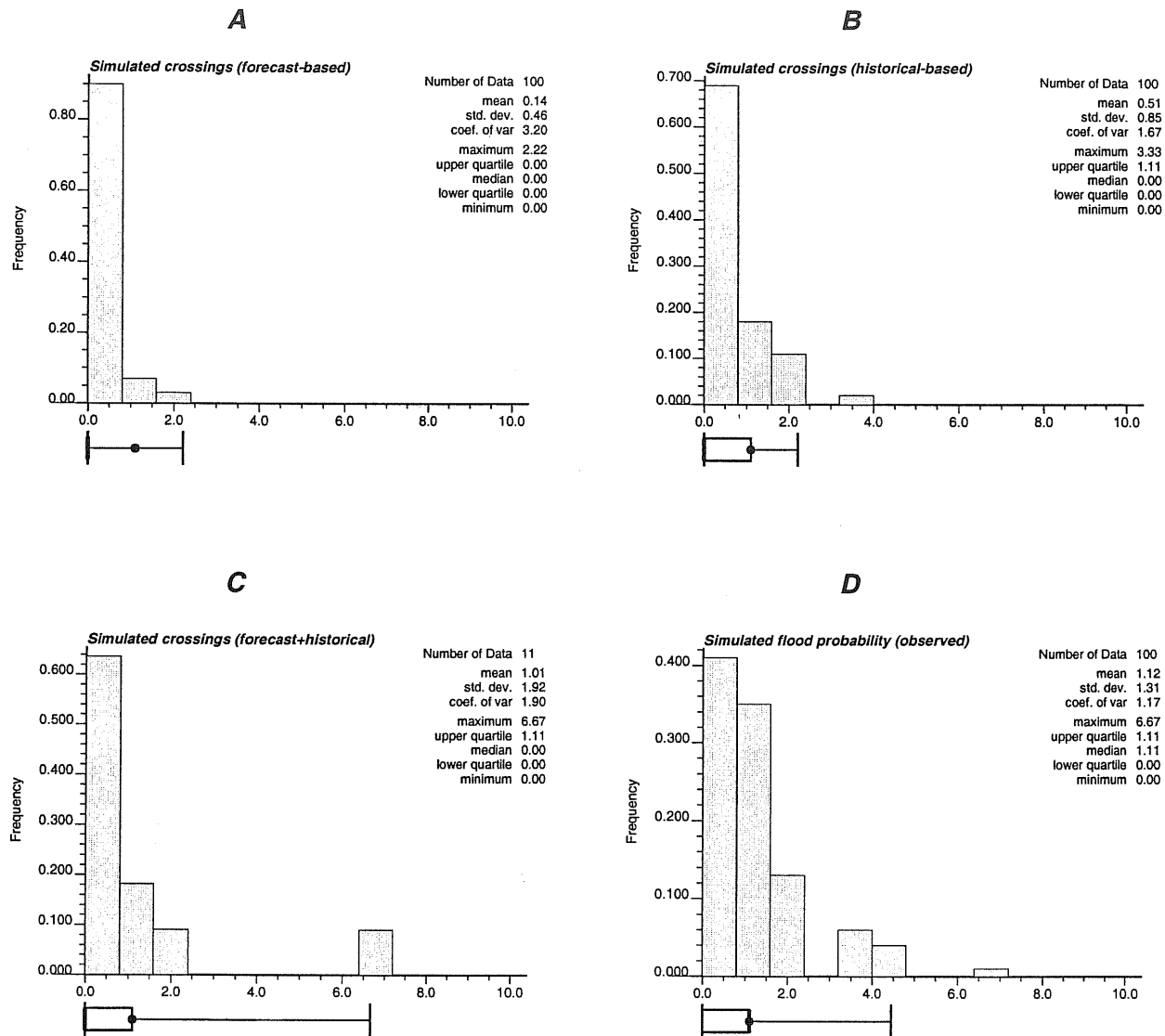
A

B

C

D

FIG. 9. Histograms of percentage of days over the forecast period for which simulated streamflow exceeds the 40 mm day^{-1} threshold value (also termed crossings), based on different parameter vectors used in the stochastic precipitation model. Such parameters were derived from (a) the forecast precipitation record, (b) the historical precipitation data, (c) both forecast and historical records, and (d) the reference precipitation at Hopland. The bullet (●) indicates the reference (TOPMODEL based) percentage of 1.11%; the corresponding forecast-based percentage is zero.

approximates the observed record in the case of correct parameter identification. Further developments for explicit handling of precipitation spatial variability will be reported in the near future.

Stochastic simulation could also be used for arriving at ensemble daily forecasts in a perturbed climate projection setting. Such daily ensemble forecasts could be used for probabilistic inference in impact assessment studies, for example, for calculating flood probabilities in a doubled carbon dioxide scenario. In such extrapolative scenarios, however, additional sources of uncertainty, for example, the impact of unknown soil moisture on streamflow, should be also investigated along with uncertainty in precipitation forcing. Parameter un-

certainty and system feedbacks, not addressed in this study, could be of critical importance in a climate change context.

Last, it should be stressed that probabilistic analysis provides a much richer input to risk-based decisions, instead of a single projected answer. Probabilistic analysis allows investigation and selection of alternative scenarios based on their corresponding environmental or societal consequences. The uncertainty propagation procedure proposed in this paper allows for such a probabilistic analysis in hydrologic impact assessment studies.

REFERENCES

Abramovitz, M., and I. A. Stegun, 1972: *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables.* 9th ed. Dover, 1046 pp.

Barnett, T., 1995: Monte Carlo climate forecasting. *J. Climate,* **8,** 1005–1022.

Beven, K. J., and M. Kirkby, 1979: A physically-based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.,* **24,** 43–69.

——, and A. Binley, 1992: The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Proc.,* **6,** 279–298.

Box, G. E. P., and G. C. Tiao, 1973: *Bayesian Inference in Statistical Analysis.* Addison-Wesley, 588 pp.

Bras, R. L., and I. Rodríguez-Iturbe, 1985: *Random Functions and Hydrology.* Addison-Wesley, 559 pp.

Caers, J. K., 2000: Adding local accuracy to direct sequential simulation. *Math. Geol.,* **32,** 815–850.

Carlin, B. P., and T. A. Louis, 2000: *Bayes and Empirical Bayes Methods for Data Analysis.* 2d ed. Chapman and Hall/CRC, 400 pp.

Chang, T. J., M. L. Kavvas, and J. W. Delleur, 1984: Daily precipitation modeling by discrete autoregressesive moving average processes. *Water Resour. Res.,* **20,** 565–580.

Chebaane, M., J. D. Salas, and D. C. Boes, 1995: Product periodic autoregressive processes for modeling intermittent monthly streamflows. *Water Resour. Res.,* **31,** 1513–1518.

Deutsch, C. V., and A. G. Journel, 1998: *GSLIB: Geostatistical Software Library and User's Guide.* 2d ed. Oxford University Press, 368 pp.

Entekhabi, D., and Coauthors, 1999: An agenda for land surface hydrology research and a call for the second international hydrological decade. *Bull. Amer. Meteor. Soc.,* **80,** 2043–2058.

Foufoula-Georgiou, E., and W. Krajewski, 1995: Recent advances in rainfall modeling, estimation, and forecasting. *Reviews of Geophysics, U.S. National Report to International Union of Geodesy and Geophysics 1991–1994,* American Geophysical Union, 1125–1137.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 1995: *Bayesian Data Analysis.* Chapman and Hall/CRC, 526 pp.

Journel, A. G., 1989: *Short Course in Geology.* Vol. 8, *Fundamentals of Geostatistics in Five Lessons,* American Geophysical Union, 40 pp.

——, 1999: Conditioning geostatistical operations to nonlinear volume averages. *Math. Geol.,* **31,** 931–953.

Katz, R. W., and M. B. Parlange, 1995: Generalizations of chain-dependent processes: Application to hourly precipitation. *Water Resour. Res.,* **31,** 1331–1341.

Kim, J., N. L. Miller, A. K. Guetter, and K. P. Georgakakos, 1998:

River flow response to precipitation and snow budget in California during the 1994/95 winter. *J. Climate,* **11,** 2376–2386.

——, ——, J. D. Farrara, and S.-Y. Hong, 2000: A seasonal precipitation and stream flow hindcast and prediction study for the 1997/98 winter season using a dynamic downscaling system. *J. Hydrometeor.,* **1,** 311–329.

Kitanidis, P. K., 1986: Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resour. Res.,* **22,** 499–507.

Krzysztofowicz, R., 1998: Probabilistic hydrometeorological forecasts: Toward a new era in operational forecasting. *Bull. Amer. Meteor. Soc.* **79,** 243–251.

——, 1999: Bayesian theory for probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.,* **35,** 2739–2750.

Lettenmaier, D. P., 1994: Application of stochastic modeling in climate change impact assessment. *Stochastic and Statistical Methods in Hydrology and Environmental Engineering,* K. W. Hipel et al., Eds., Time Series Analysis in Hydrology and Environmental Engineering, Vol. 3, Kluwer Academic, 3–17.

Leung, L. R., M. S. Wigmosta, S. J. Ghan, D. J. Epstein, and L. W. Vail, 1996: Application of a subgrid orographic precipitation/surface hydrology scheme to a mountain watershed. *J. Geophys. Res.,* **101,** 12 803–12 817.

——, A. F. Hamlet, D. P. Lettenmaier, and A. Kumar, 1999: Simulations of the ENSO hydroclimate signals in the Pacific Northwest Columbia River basin. *Bull. Amer. Meteor. Soc.,* **80,** 2313–2329.

Mardia, K. V., J. T. Kent, and J. M. Bibby, 1979: *Multivariate Analysis.* Academic Press, 518 pp.

Miller, N. L., and J. Kim, 1996: Numerical prediction of precipitation and river flow over the Russian River watershed during the January 1995 California storms. *Bull. Amer. Meteor. Soc.,* **77,** 101–105.

——, ——, R. K. Hartman, and J. D. Farrara, 1999: Downscaled climate and streamflow study of the southwestern United States. *J. Amer. Water Resour. Assoc.,* **35,** 1525–1538.

Ripley, B., 1987: *Stochastic Simulation.* John Wiley and Sons, 256 pp.

Salas, J. D., 1993: Analysis and modeling of hydrologic time series. *Handbook of Hydrology,* D. R. Maidment, Ed., McGraw-Hill, 19.1–19.71.

Tanner, M. A., 1996: *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* Springer-Verlag, 207 pp.

Valdés, J. B., I. Rodríguez-Iturbe, and G. J., Vicens, 1977: Bayesian generation of synthetic streamflows, 2: The multivariate case. *Water Resour. Res.,* **13,** 291–295.

Valencia, R. D., and J. C. Schaake, 1973: Dissaggregation processes in stochastic hydrology. *Water Resour. Res.,* **9,** 580–585.

Wilby, R. L., T. M. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.,* **34,** 2995–3008.

Wilks, D. S., 1992: Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change,* **22,** 67–84.

——, 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

Woolhiser, D. A., 1992: Modeling daily precipitation—progress and problems. *Statistics in Environmental and Earth Sciences,* A. T. Walden and P. Guttorp, Eds., Edward Arnold, 71–89.