

A Bayesian Processor of Uncertainty for Precipitation Forecasting Using Multiple Predictors and Censoring

PAOLO REGGIANI AND OLEKSIY BOYKO

Department of Civil Engineering, University of Siegen, Siegen, Germany

(Manuscript received 12 March 2019, in final form 18 July 2019)

ABSTRACT

A Bayesian processor of uncertainty for numerical precipitation forecasts is presented. The predictive density is estimated on the basis of normalized variates, the use of censored distributions, and the implementation of a parameter-parsimonious and computationally efficient processor that is applicable in operational settings. The structure of the processor is sufficiently generic to handle mixed binary-continuous random processes such as intermittent rainfall (and similarly ephemeral river flows), and an arbitrary number of predictors. First, predictors and observations, the parent data sample, are mapped into standard Gaussian variates, obtaining a nonparametric approximately multivariate normal distribution (MVND) that is considered censored for days with no precipitation. To convert the Gaussian binary-continuous multivariate precipitation process into a continuous one, the parent sample is augmented into the negative range through Bayesian imputation by Gibbs sampling, recovering the true, a priori unknown variance–covariance structure of the full uncensored sample. The dependency among marginal distributions of observations and predictions is hereby assumed multivariate normal, for which closed-form expressions of conditional densities exist. These are then mapped back into the variable space of provenience to yield the predictive density. The processor is applied to a well-monitored study area in Switzerland. Standard forecast performance evaluation and verification metrics are employed to set the approach into perspective against Bayesian model averaging (BMA).

1. Introduction


A wide series of forecasting applications require forcing hydrological or reservoir models by means of numerical weather predictions. Examples include river stage and flow forecasting, coastal flood forecasting, as well as irrigation or reservoir operations. Predictions of meteorological forcing variables are inherently uncertain due to the internal structure of specific atmospheric models, the high nonlinearity of the underlying physical processes, as well as the selection and propagation of initial and boundary conditions. The forecast uncertainty is assessed in terms of a probability density function (PDF) of the predictand, conditional on forecast series of atmospheric variables that serve as predictors. This conditional density function is referred to as predictive density (Krzysztofowicz 1999; Hamill

and Whitaker 2006) and must be provided to forecast end users:

$$f(\mathbf{x}|\hat{\mathbf{X}} = \hat{\mathbf{x}}), \quad (1)$$

with $\hat{\mathbf{x}}$ a realization of the multidimensional random vector of predictors $\hat{\mathbf{X}}$ at n time steps and \mathbf{x} a realization of the random process \mathbf{X} , say precipitation, that we expect to observe at the same future times of $\hat{\mathbf{x}}$. Predictors can include variables, such as forecasted precipitation (Seo et al. 2000; Vrac and Naveau 2007; Sloughter et al. 2007), atmospheric state variables like pressure (Bárdossy and Plate 1992), precipitable water, vorticity, or mean and variance of ensemble outputs (Hamill and Whitaker 2006; Scheuerer and Hamill 2015) of such state variables.

Unlike temperature, wind or humidity, which are envisaged as continuous random processes, precipitation is intermittent. Stochastically, the precipitation process is described as a mixture of binary and continuous variates, as described in appendix A. The mixed binary-continuous structure poses particular

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Paolo Reggiani, paolo.reggiani@uni-siegen.de

challenges when estimating the predictive density in (1) as it requires handling one-sided truncated distributions. Sloughter et al. (2007) applied Bayesian model averaging (BMA) (Raftery et al. 2005) in estimating the density in (1), and respectively (A3), and its moments:

$$\begin{aligned} E(\mathbf{X}|\hat{\mathbf{X}} = \hat{\mathbf{x}}) \\ \text{VAR}(\mathbf{X}|\hat{\mathbf{X}} = \hat{\mathbf{x}}). \end{aligned} \quad (2)$$

BMA is applied by mixing independent univariate PDFs of precipitation, conditioned on respective predictors, through linear weighting. The conditional probability of precipitation (PoP) \hat{v} in (A3) is chosen ad hoc following Hamill et al. (2004) as parametric logistic regression by using a power-transformed precipitation depth forecast as predictor. Precipitation depth is assumed gamma distributed (Wilks 1990). The linear BMA weights are determined by log-likelihood maximization.

Bárdossy and Pegram (2009) use copulas to obtain multivariate probability distributions of precipitation among gauges of a ground station network. Following Sklar's (Sklar 1959) theorem unique copulas require continuous marginal distributions for their construction, which poses difficulties when handling binary-continuous processes such as precipitation. A wide base of copula forms are available for modeling purposes. One major drawback, however, is the high number of parameters if parametric marginal distributions are used. The parameters need to be estimated through maximum-likelihood optimization. We will show that the approach proposed here is akin to modeling multivariate dependence using a nonparametric Gaussian copula.

In a similar context Herr and Krzysztofowicz (2005) proposed a closed-form approach of bivariate precipitation modeling. The mixed binary-continuous precipitation process, observed at two sites, is mapped to standard normal variates using the nonparametric normal quantile transform (NQT). The bivariate dependence is modeled in terms of the mutually conditional PoP, two marginal and two conditional distributions and the covariance as parameter, 8 elements in total. Assuming a bivariate standard normal dependence is equivalent to using a Gaussian copula model. Albeit rigorous verification of the model, the extension of the bivariate case to multivariate is impractical because of the prohibitive number of conditional PoP combinations and conditional/marginal distributions (see appendix A). For instance, for two predictors and one predictand, the number of necessary parameters is 10 and 6 univariate distributions.

For three and more predictors, the number of parameters grows factorially.

The above difficulties of proposing suitable parametric multivariate distribution models favors random sampling. To avoid handling mixed binary-continuous distribution structures with precipitation (and similarly, intermittent river flows), variates can be assumed as censored. Censored distributions describe random processes, in which measurements are cut off beyond a critical threshold, while the amount of censoring points is known. We note that censoring is different from truncation, as the latter by definition excludes values beyond the truncation threshold. As a result, mean and variance of a censored and a truncated distribution differ. Areas of application for censored distributions include life sciences (Sorensen et al. 1998) and system failure analysis (Kalbfleisch and Prentice 1980). The parameters and conditional distributions for censored distribution can be inferred by means of data augmentation through Markov chain Monte Carlo (MCMC) sampling. In hydrometeorology, censored distributions have been used in different contexts for postprocessing raw precipitation ensemble forecasts (Scheuerer and Hamill 2015), or data series, that have been power transformed to approximate a normal distribution. Bárdossy and Plate (1992) applied a simple parametric power transformation to precipitation data, while Frost et al. (2007) and Wang et al. (2011) used a modified Box-Cox and Yeo-Johnson power transform to normalize river discharge data. All reported applications are limited to a small number of predictors due to heavy parameterization.

Here we propose an alternative approach by specifying the predictive density (1) through a mixture of concepts adopted in the previous approaches. The proposal rests on the model-conditional processor (MCP) proposed by Todini (2008) for probabilistic flow forecasting. Recently Reggiani et al. (2019) used the MCP to process monthly precipitation reanalysis in poorly gauged basins in Pakistan. So far the MCP has been applied to the derivation of predictive densities for continuous variates such as river stage and discharge (Coccia and Todini 2011), monthly average precipitation and temperature (Reggiani et al. 2016). Now we extend the approach to multivariate mixed binary-continuous variates, primarily daily precipitation. The principal advantages of the approach can be summarized as follows:

- 1) The processor is parameter-parsimonious and thus applicable to large multivariate problems.
- 2) It does not require the use of ad hoc assumptions on the choice of PoP or parametric precipitation

depth CDFs due to the direct use of empirical distributions obtained from observed data and the use of nonparametric transformations to standard normal variates.

- 3) It is extensible to probabilistic forecasting for arbitrarily located ground stations and an arbitrary number of predictors, suggesting applicability to ensemble forecasting.
- 4) It is computationally efficient and thus optimized for operational use.

We introduce the structure of the processor and show its properties and performance verification on a spatially limited example, while more extended applications will be given in a sequel work. As such, the current work should be envisaged as proof of concept rather than an operational application. The text is structured as follows: in [section 2](#) we introduce the theory, in [section 3](#) we describe data and processor application, in [section 4](#) we describe processor execution and results, in [section 5](#) we provide a discussion, and in [section 6](#) we present the conclusions. Implementation details are given in the appendices.

2. Methods

a. Processor model

Given $F(\mathbf{X} \leq \mathbf{x})$ and $G(\hat{\mathbf{X}} \leq \hat{\mathbf{x}})$ are marginal distributions, as well as monotonous and continuous functions of predictand and predictor introduced in (1), their Gaussian transforms are $\mathbf{w} = \Phi^{-1}[F(\mathbf{x})]$ and $\mathbf{z} = \Phi^{-1}[G(\hat{\mathbf{x}})]$ with Φ^{-1} the inverse of the standard normal distribution. As indicated in [appendix B](#), any variate can be transformed into a standard-normal one by means of the nonparametric quantile transform (NQT) ([Moran 1970](#)). The Gaussian variates are random vectors that are structured as follows:

$$\begin{aligned} \mathbf{W} &= [W_1, \dots, W_n]^T \\ \mathbf{Z} &= [(Z_{11}, \dots, Z_{n1}), \dots, (Z_{1m}, \dots, Z_{nm})]^T, \end{aligned} \quad (3)$$

where $1, \dots, n$ are the number of time steps and m is the number of predictors. Next, we assume that the standard-normal predictand \mathbf{W} can be related to the predictors \mathbf{Z} through a linear model:

$$\mathbf{W} = \mathbf{A} + \mathbf{B} \cdot \mathbf{Z} + \mathbf{\Omega}, \quad (4)$$

where \mathbf{B} is a $1 \times m$ vector consisting of the multilinear regression coefficients of the model, \mathbf{A} is a $n \times 1$ vector of intercepts, and $\mathbf{\Omega} \sim N(0, \sigma_{\mathbf{z}, \mathbf{\Omega}}^2)$ is a Gaussian noise, stochastically independent of \mathbf{W} and \mathbf{Z} . The conditional mean and variance of \mathbf{W} are given as follows:

$$\begin{aligned} E(\mathbf{W}|\mathbf{Z} = \mathbf{z}) &= \mathbf{A} + \mathbf{B} \cdot \mathbf{z} \\ \text{Var}(\mathbf{W}|\mathbf{Z} = \mathbf{z}) &= \sigma_{\mathbf{z}, \mathbf{\Omega}}^2. \end{aligned} \quad (5)$$

We note that $\sigma_{\mathbf{z}, \mathbf{\Omega}}^2$ varies with \mathbf{z} for heteroscedastic dependency and is constant for homoscedastic structures. Prior to any verification for homoscedasticity, we retain the dependency on \mathbf{z} in the notation. As will be shown in [section 3e](#), the data used in this study are essentially homoscedastic and thus independent of \mathbf{z} . In our case the predictand \mathbf{x} with NQT transform \mathbf{w} is an observed point precipitation measurement series, spatially averaged over a square grid cell (analysis cell), while the predictors $\hat{\mathbf{x}}$ with NQT transform \mathbf{z} are corresponding forecast series provided by a weather model at 8 grid cells surrounding the analysis cell and at the cell overlapping the latter (see [Fig. 1](#)), 9 cells in total.

ASSUMPTION

The joint $n \times (m + 1)$ -sized sample of NQT-transformed observations and predictions (\mathbf{w}, \mathbf{z}) is multivariate normal distributed (MVND):

$$(\mathbf{W}, \mathbf{Z}) \sim N_{(m+1)}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (6)$$

with density

$$\phi(\mathbf{w}, \mathbf{z}) = \frac{\exp\left\{-\frac{1}{2}[(\mathbf{w}, \mathbf{z}) - \boldsymbol{\mu}]^T \boldsymbol{\Sigma}^{-1}[(\mathbf{w}, \mathbf{z}) - \boldsymbol{\mu}]\right\}}{\sqrt{(2\pi)^{m+1} |\boldsymbol{\Sigma}|}}, \quad (7)$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the joint sample (\mathbf{w}, \mathbf{z}) and $\boldsymbol{\mu} = E(\mathbf{w}, \mathbf{z})$ the $(m + 1)$ -sized row vector of sample means, equal to zero for standard-normal distributed variables. We note that assuming the joint distribution multivariate normal is equivalent to choosing a Gaussian copula dependence for the normal marginal distributions F and G . By virtue of the inherent properties of multivariate normal distributions ([Mardia et al. 1979](#)) closed-form conditional densities can be obtained from (7):

$$\phi(\mathbf{w}|\mathbf{z}) = \frac{\phi(\mathbf{w}, \mathbf{z})}{\phi(\mathbf{z})} = \frac{\exp\left\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu}_{\mathbf{w}|\mathbf{z}})^2 / \sigma_{\mathbf{w}|\mathbf{z}}^2\right\}}{\sqrt{2\pi\sigma_{\mathbf{w}|\mathbf{z}}^2}}, \quad (8)$$

with $\phi(\mathbf{z})$ the marginal density of (7). Equation (8) represents a family of predictive densities of type in (1) in the standard normal space with conditional mean and variance:

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{w}|\mathbf{z}} &= \boldsymbol{\Sigma}_{\mathbf{wz}} \boldsymbol{\Sigma}_{\mathbf{zz}}^{-1} \mathbf{z} \\ \sigma_{\mathbf{w}|\mathbf{z}}^2 &= 1 - \boldsymbol{\Sigma}_{\mathbf{wz}} \boldsymbol{\Sigma}_{\mathbf{zz}}^{-1} \boldsymbol{\Sigma}_{\mathbf{zw}}, \end{aligned} \quad (9)$$

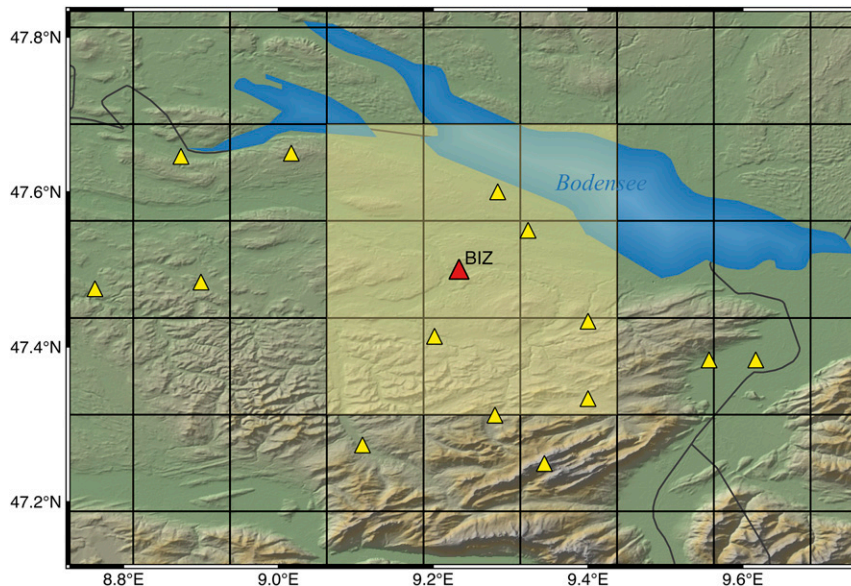


FIG. 1. The ERA-Interim weather model grid and analysis window including nine cells evidenced by the shadowed area. The triangles represent 15 observing stations including Bischofszell (BIZ) in the central analysis cell. Observed precipitation has been mapped from points to cell averages by block kriging.

where Σ_{zz} is the $m \times m$ covariance matrix of \mathbf{Z} and Σ_{wz} the $1 \times m$ covariance vector of \mathbf{Z} and \mathbf{W} . From the equivalence between the model in (5) and (9) it follows that $\mathbf{A} = 0$, $\mathbf{B} = \Sigma_{wz}\Sigma_{zz}^{-1}$, and $\sigma_{z,\Omega}^2 = 1 - \Sigma_{wz}\Sigma_{zz}^{-1}\Sigma_{zw}$, a constant value. In other words, in the standard normal space the multilinear model in (5) crosses the origin and has residuals with constant variance over the entire range of \mathbf{z} . The validity of this model must undergo testing as discussed in section 3e.

The predictive density in (8) is a well-calibrated uncertainty assessor (Alpert and Raiffa 1982; Krzysztofowicz 1999), as statistics of the processed precipitation (mean and variance) match those of retrospective observations. It is easy to verify that the processor, which produces a density of \mathbf{w} drawn from the family of conditional densities $\{\phi(\cdot|\mathbf{z}); \forall \mathbf{z}\}$, satisfies the condition:

$$E_z[\phi(\mathbf{w}|\mathbf{z})] = \phi(\mathbf{w}), \quad (10)$$

where the expectation, taken over all \mathbf{z} , is equivalent to marginalization, and $\phi \equiv N(0, 1)$ is the prior climatic density of standard normal observations. The processor is moreover self-calibrating, meaning that it delivers a well-calibrated output also if the predictors output by the weather forecast model are not well calibrated. The Gaussian predictive density in (8) needs to be finally mapped back into the predictive density of the

binary-continuous process in the real space, as described in section 4a.

b. Censoring and nonignorable missing data

One way of handling precipitation, an intermittent random process, is transforming it into a continuous one. This is achieved by assuming zero precipitation as a nonobserved value that has been censored below the zero-precipitation threshold, respectively, its corresponding value in the standard normal space. The only information known about the censored series is the number r of missing values. The joint sample (\mathbf{w}, \mathbf{z}) is horizontally subdivided into two parts, the observations \mathbf{y}^o and the censored values \mathbf{y}^c :

$$\mathbf{Y} = (\mathbf{w}, \mathbf{z}) = \begin{bmatrix} y_{1,1}^o, \dots, y_{1,m+1-r_1}^o, y_{1,1}^c, \dots, y_{1,r_1}^c \\ \dots \\ y_{n,1}^o, \dots, y_{n,m+1-r_n}^o, y_{n,1}^c, \dots, y_{n,r_n}^c \end{bmatrix} = (y_{j,k}), \quad (11)$$

with \mathbf{Y} an $n \times (m+1)$ matrix obtained by horizontally concatenating the NQT-transformed predictand \mathbf{w} and m predictors \mathbf{z} defined in (3). For a given row vector $\mathbf{y} = (\mathbf{y}^o, \mathbf{y}^c)$ of \mathbf{Y} , the subvectors \mathbf{y}^o and \mathbf{y}^c can have different lengths according to the censoring threshold value c for that column. For reasons of notational simplicity we omit the index j when addressing a specific row. We also note that the sample contains

rows, where all data are censored because at that particular time step observations and predictions are all zero, thus $r = m + 1$.

To fill (or impute) the gap of unknown missing values in each row of the sample, we apply a random sampling technique, which preserves the mean and the variance-covariance structure of the whole sample, including the yet unknown censored values. We note that these moments are different from those of the same but truncated sample, where in contrast to censoring, no values exist beyond the truncation threshold. Depending on the method used, the sampling is known as data augmentation (Tanner and Wong 1987) or imputation (Little and Rubin 2002).

First we denote with $f(\mathbf{y}|\theta) = f(\mathbf{y}^o, \mathbf{y}^c|\theta)$ the joint probability density function of \mathbf{y} , conditional on the parameter θ . For a Gaussian model $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In particular cases the mechanism leading to missing data through censoring can be ignored, and the marginal probability density of \mathbf{y}^o is thus obtained by integrating out \mathbf{y}^c :

$$f(\mathbf{y}^o|\theta) = \int f(\mathbf{y}^o, \mathbf{y}^c|\theta) d\mathbf{y}^c. \quad (12)$$

In our case, however, the censoring mechanism is perfectly known and thus nonignorable. The model in (12) needs therefore to be enhanced. To this end we introduce a missing data mechanism formulated in terms of an $n \times (m + 1)$ -dimensional matrix \mathbf{M} , which a priori is a variate describing the missingness pattern of \mathbf{y}^c . In our case missing data arise from censoring the column k of \mathbf{Y} below the precipitation threshold c_k , so that only values larger than c_k are recorded. The members of \mathbf{M} are described by the conditional PDF of the binary-discrete variate $M_{j,k}$:

$$f(M_{j,k}|y_{j,k}) \quad (13)$$

taking on the following values:

$$f(M_{j,k}|y_{j,k}) = \begin{cases} 1, & \text{if } M_{j,k} = 1 \text{ and } y_{j,k} \leq c_k \text{ or} \\ & M_{j,k} = 0 \text{ and } y_{j,k} > c_k \forall j, k. \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

We restate (12) by explicitly carrying the dependence on \mathbf{M} :

$$f(\mathbf{y}^o, \mathbf{M}|\theta) = \int f(\mathbf{M}|\mathbf{y})f(\mathbf{y}|\theta) d\mathbf{y}^c \quad \mathbf{y} = \{\mathbf{y}^o, \mathbf{y}^c\}. \quad (15)$$

If the data censoring mechanism is independent of \mathbf{y}^c , data are said to be missing at random (MAR) (Little

and Rubin 2002). Contrarily and as in our case, data are missing not at random (MNAR).

We note that the binary variate \mathbf{M} introduces an additional degree of randomization, which accounts for the probability of a draw being at or below threshold. This uncertainty is absent for MAR data and for truncated distributions, which by definition are confined inside the truncation boundaries. The additional randomization leads to inflation of mean and variance of the posterior distribution.

The parameters can be estimated by applying expectation maximization (EM) or the Newton-Raphson method for maximum likelihood estimation (MLE) on (15). Both approaches can become computationally demanding if dealing with complex missingness pattern and large multivariate MNAR data. An alternative approach is Bayesian imputation.

c. Bayesian imputation

We introduce Bayesian imputation by first considering the simple case of a semicomplete sample of data \mathbf{y}^o with sparsely missing observations, effectively a MAR situation with ignorable \mathbf{M} . We apply Bayes's theorem and infer the posterior distribution on the parameter θ :

$$f(\theta|\mathbf{y}^o) \propto \mathcal{L}(\theta|\mathbf{y}^o)p(\theta), \quad (16)$$

where $p(\theta)$ is the prior distribution of θ , while the likelihood function is defined as a family of density functions proportional to the posterior density:

$$\mathcal{L}(\theta|\mathbf{y}^o) \propto f(\mathbf{y}^o|\theta), \quad (17)$$

with the proportionality factor consisting of a normalizing constant independent of θ :

$$f(\theta|\mathbf{y}^o) = \frac{f(\mathbf{y}^o|\theta)p(\theta)}{\int f(\mathbf{y}^o|\theta)p(\theta) d\theta}. \quad (18)$$

The integral at the denominator can be evaluated analytically only in particular cases. Otherwise one needs to sample directly from the posterior distribution.

In our problem of interest, however, we need to state the complete-case MNAR Bayesian problem, which includes the missing data in addition to the missing-data mechanism \mathbf{M} specified by (14). Bayesian inference leads to statement of the joint posterior distribution for θ and \mathbf{y}^c , conditional on known information \mathbf{M} and \mathbf{y}^o :

$$f(\theta, \mathbf{y}^c|\mathbf{y}^o, \mathbf{M}). \quad (19)$$

This distribution can be factored by applying the chain rule of probabilities:

$$f(\theta, \mathbf{y}^c | \mathbf{y}^o, \mathbf{M}) = f(\theta | \mathbf{y}, \mathbf{M}) f(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M}); \quad \mathbf{y} = \{\mathbf{y}^o, \mathbf{y}^c\}. \quad (20)$$

Marginalization with respect to \mathbf{y}^c yields the posterior predictive density of the parameter θ :

$$f(\theta | \mathbf{y}^o, \mathbf{M}) = \int f(\theta | \mathbf{y}, \mathbf{M}) f(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M}) d\mathbf{y}^c. \quad (21)$$

The complete-data posterior $f(\theta | \mathbf{y}, \mathbf{M})$ is used for the complete-case analysis, as will become clear below. Next, we perform an alternative factorization of (19):

$$f(\theta, \mathbf{y}^c | \mathbf{y}^o, \mathbf{M}) = f(\mathbf{y}^c | \theta, \mathbf{y}^o, \mathbf{M}) f(\theta | \mathbf{y}^o, \mathbf{M}) \quad (22)$$

and marginalize out θ :

$$f(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M}) = \int f(\mathbf{y}^c | \theta, \mathbf{y}^o, \mathbf{M}) f(\theta | \mathbf{y}^o, \mathbf{M}) d\theta \quad (23)$$

to obtain the posterior predictive density on \mathbf{y}^c . From this density we sample $\mathbf{y}^{c,(t)}$, which is used in the complete-data posterior $f(\theta | \mathbf{y}, \mathbf{M})$ to draw $\theta^{(t)}$ in successive Markov chain Monte Carlo (MCMC) sampling steps. It can be shown that the two sequences approach the joint limiting density in (19) for $t \rightarrow \infty$. The subsequence $\theta^{(t)}$, marginalized with respect to \mathbf{y}^c , approximates the integral in (21) with limiting density $f(\theta | \mathbf{y}^o, \mathbf{M})$, while the imputed subsequence $\mathbf{y}^{c,(t)}$ marginalized with respect to θ , approximates (23) with limiting density $f(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M})$.

In Bayesian terms, a family of likelihood functions, which are proportional to the densities $\mathcal{L}(\theta | \mathbf{y}, \mathbf{M}) \propto f(\mathbf{y}, \mathbf{M} | \theta)$ and $\mathcal{L}(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M}, \theta) \propto f(\theta, \mathbf{y}^o, \mathbf{M} | \mathbf{y}^c)$, are used in successive inference steps to revise prior into posterior information:

$$\mathbf{y}^{c,(t)} \sim f(\mathbf{y}^c | \theta^{(t-1)}, \mathbf{y}^o, \mathbf{M}) \propto p(\mathbf{y}^c) \mathcal{L}(\mathbf{y}^c | \theta^{(t-1)}, \mathbf{y}^o, \mathbf{M}) \quad (24)$$

In practice we draw $\mathbf{y}^{c,(t)}$ and $\theta^{(t)}$ of the random sequence directly from the posterior distribution, a process also known as imputation or data augmentation. Because the sequential random sampling process is ergodic, convergence toward the expectation of the joint posterior distributions (the ensemble mean) given by integrals (21) and (23) is assured:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(\theta | \mathbf{y}^o, \mathbf{y}^{c,i}, \mathbf{M}) &\rightarrow f(\theta | \mathbf{y}^o, \mathbf{M}) \\ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(\mathbf{y}^c | \theta^i, \mathbf{y}^o, \mathbf{M}) &\rightarrow f(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M}). \end{aligned} \quad (25)$$

Effectively the conditional distributions of the posteriors (21) and (23) can be replaced by their summaries, the conditional mean:

$$\begin{aligned} E(\theta | \mathbf{y}^o, \mathbf{M}) &= E[E(\theta | \mathbf{y}^o, \mathbf{y}^c, \mathbf{M}) | \mathbf{y}^o, \mathbf{M}] \\ E(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M}) &= E[E(\mathbf{y}^c | \theta, \mathbf{y}^o, \mathbf{M}) | \mathbf{y}^o, \mathbf{M}]. \end{aligned} \quad (26)$$

and variance:

$$\begin{aligned} \text{Var}(\theta | \mathbf{y}^o, \mathbf{M}) &= E[\text{Var}(\theta | \mathbf{y}^o, \mathbf{y}^c, \mathbf{M}) | \mathbf{y}^o, \mathbf{M}] \\ &\quad + \text{Var}[E(\theta | \mathbf{y}^o, \mathbf{y}^c, \mathbf{M}) | \mathbf{y}^o, \mathbf{M}] \\ \text{Var}(\mathbf{y}^c | \mathbf{y}^o, \mathbf{M}) &= E[\text{Var}(\mathbf{y}^c | \theta, \mathbf{y}^o, \mathbf{M}) | \mathbf{y}^o, \mathbf{M}] \\ &\quad + \text{Var}[E(\mathbf{y}^c | \theta, \mathbf{y}^o, \mathbf{M}) | \mathbf{y}^o, \mathbf{M}], \end{aligned} \quad (27)$$

which can both be approximated as sums analogous to (25) for sufficiently large number of draws n (Little and Rubin 2002).

d. MCMC sampling

We use MCMC simulation to generate a large number of sample values from the two distributions in (24) for \mathbf{y}^c and the parameters $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, respectively, and approximate the summaries $E[\bullet]$ and $\text{Var}[\bullet]$ of interest directly from the sample. The MCMC simulator is implemented as nested Gibbs sampler (Geman and Geman 1984) with an “inner” sampler nested in an “outer” sampling cycle. The outer Gibbs sampler sweeps over n time steps of the $m + 1$ -dimensional sample, while the inner sampler is used to draw from a truncated multinormal distribution at each time step, as recovery of the censored observations requires simulation from the corresponding truncated normal distribution (Kotecha and Djurić 1999). Gibbs sampling in this latter case works more efficiently than rejection sampling from the truncated MVND. Details of the nested Gibbs sampler implementation are given in appendix C.

After a series of sampling steps, during which the MCMC process loses track about the arbitrarily chosen set of initial parameter values (the burn-in period), the values sampled at each iteration represent a draw from the posterior distribution, and the statistics in (26) and (27) can be computed to a degree of approximation, which depends on of the number of sampled values.

3. Application

a. Study site

A well-monitored area in the northern part of Switzerland bordering on Lake Constance and depicted in Fig. 1 was selected as study site. The area is served by 15 meteorological stations operated by Meteo Swiss

with an hourly recording interval. The figure also shows the grid of the ERA-Interim weather forecasting model at $0.125^\circ \times 0.125^\circ$ spatial resolution. We note that the chosen resolution corresponding to the Gaussian N640 Grid is smaller than the one of the native $0.7^\circ \times 0.7^\circ$ N128 Gaussian grid. The downsizing was performed with the aid of the Meteorological Archival and Retrieval System (MARS) of ECMWF. We select the square containing station Bischofszell (BIZ) as analysis cell and the 3×3 beige-shadowed window centered on the analysis cell as spatial processing region. Analyzed precipitation reforecasts for the nine cells at daily time step are used as predictors to calibrate and validate the processor. We emphasize that we have used the forecasting field contained in the ERA-Interim dataset, initialized from analyses at 0000 and 1200 UTC (Berrisford et al. 2011) and not a genuine historical forecast or reforecast. However, the procedure outlined here remains independent of such choice and can be applied to assess the uncertainty of real-time forecasts as well as reanalyses in exactly the same way.

A time window of observations and forecasts covering a continuous period of 36 years starting on 1 January 1979 to 31 December 2015 is chosen. The reforecasts are aggregated from 3-hourly to daily time steps. The whole sample includes a continuous series of data with a total number of 13 514 time steps. Precipitation lower than 0.5 mm day^{-1} is considered a nonevent. In summary we study a $m + 1 = 10$ -dimensional multivariate problem including 9 predictors and a single predictand of cell-averaged daily observations at analysis cell BIZ.

b. Block kriging

To use predictors provided at the scale of a model cell, precipitation needs to be upscaled from the hourly point measurements at individual stations to daily values at the scale of the analysis cell. For this purpose block kriging is used, a geostatistical method, in which the spatial correlation structure of the station records is represented by an empirical semivariogram, which is modeled through parametric functions. The semivariogram is time dependent and thus needs to be refitted periodically. We chose from four different semivariogram models, which were selected on the basis of optimal weighted least squares fitting (Cressie 1985). The optimal parameters are found by means of the Newton–Raphson method by minimizing the least squares error used as cost function. Alternatively one can consider the variogram parameters as random variables, which are optimized by maximum likelihood estimation (MLE) (Todini and Pellegrini 1999).

c. Normalization of variables

Next we use the NQT in [appendix B](#) to map the predictand and predictor variates \mathbf{X} and $\tilde{\mathbf{X}}$ with marginal distributions F and G into standard normal variates \mathbf{W} and \mathbf{Z} . Given that precipitation is intermittent, we consider observations on dry days as censored, whereby the censored data are supposed to belong to the fictive negative precipitation range. This enables one to associate probabilities with each single value in the record. These are matched with standard normal probabilities. After application of the inverse normal CDF Φ^{-1} a specific standard normal variate value is associated with the cutoff value of observed or forecasted precipitation set at 0.5 mm day^{-1} , which defines the corresponding Gaussian censoring thresholds for each series, given that the number of dry and wet days in each series is different.

d. Missing data imputation

The NQT transformed censored series are combined into a joint distribution, which is assumed a censored MVND [see (6)] with density in (7). Of course this assumption must be tested. Such testing will be performed after the censored sample has been extended by imputation using Gibbs sampling. Bayesian imputation leads to a $m + 1$ -dimensional complete MVN sample, including imputed values \mathbf{y}^c , by fully preserving the parameters structure $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the uncensored parent sample. The Gibbs sampler is first tested on a synthetic 10-dimensional, perfectly MVN distributed sample that is generated by random draws from the MVND with known variance-covariance matrix $\boldsymbol{\Sigma}$ and zero mean $\boldsymbol{\mu}$. The complete synthetic sample is censored using the thresholds vector \mathbf{c} calculated from the observation and forecast sample of the Swiss study site. By applying Gibbs sampling as described in [appendix C](#), the censored part \mathbf{y}^c of the sample is retrieved by imputation from the remainder of the parent sample. The variance-covariance matrix of the reconstructed sample ($\mathbf{y}^c, \mathbf{y}^o$) was compared against $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the original parent sample prior to censoring. Both parameters turned out to be essentially equal, confirming the correct convergence of the sampler through the imputation process. [Figure 2](#) shows a bivariate Gaussian synthetic sample, which was drawn, censored and then reconstructed by imputation. When working with a real-world sample of observed and forecasted precipitation assumed as censored, the parameters of the uncensored parent sample are unknown and must be iteratively recovered through estimates from the reconstructed sample. A posterior verification by comparison, as in the synthetic case, is in this case not possible.

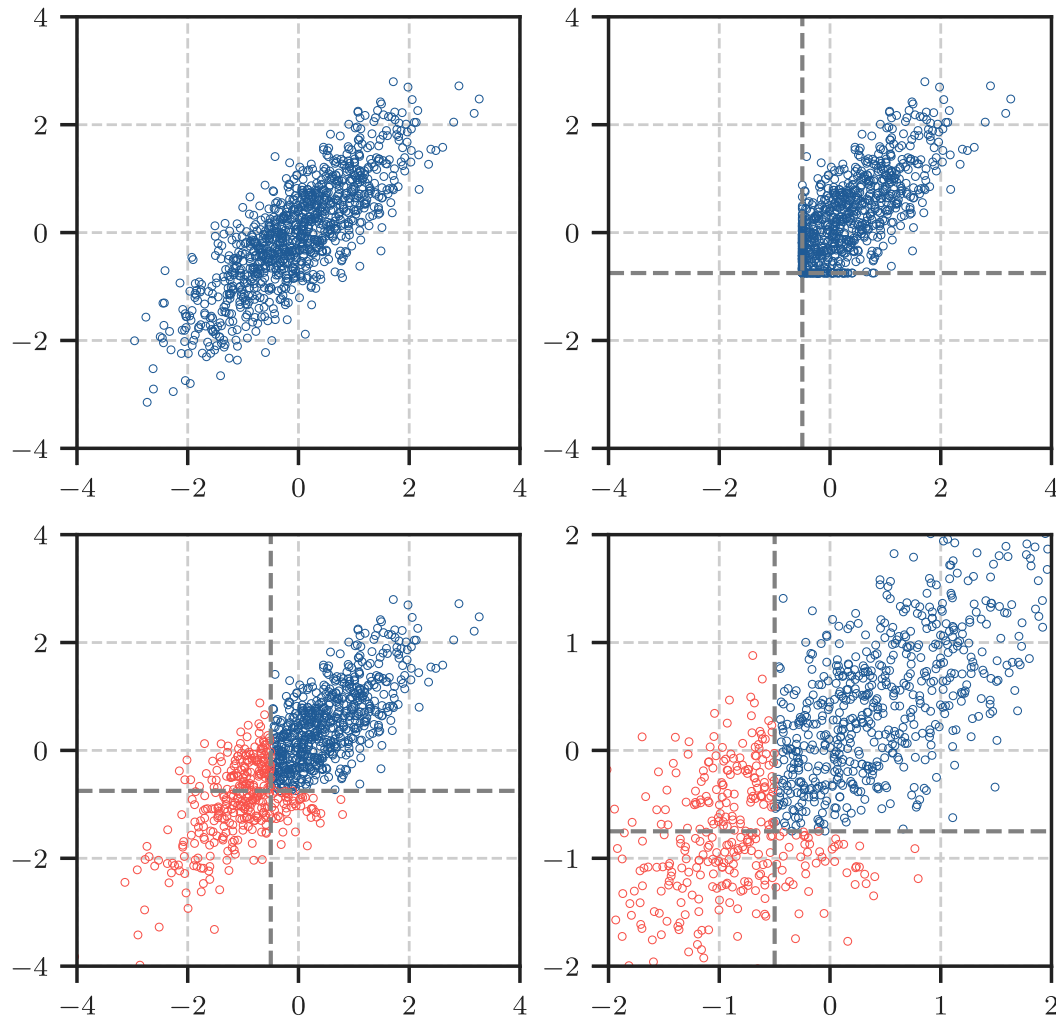


FIG. 2. Bayesian imputation for a synthetic bivariate normal sample, 10^3 draws. (top left) Original sample drawn from $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = 0$, $\sigma_1^2 = \sigma_2^2 = 1.0$, $\sigma_{12} = 0.8$. (top right) Bivariate distribution truncated at $\mathbf{c} = (-0.5, -0.75)$. (bottom left) Truncated sample with the red part retrieved by imputation, and (bottom right) a zoom in on the transition zone between parent and imputed sample.

Next we proceed with testing the Gibbs sampler on the $10 \times 13\,514$ 36-year daily dataset for the study site, including 9 predictors and one series of observations. The standard-normal values of the zero-precipitation thresholds are calculated and the Gibbs sampler applied to impute the fictitious subthreshold data values. This process requires particular attention due to the high covariance among predictor series (forecasted precipitation series from adjacent weather model grid cells in Fig. 1), which lead to poor mixing properties in the Gibbs sampling (Raftery and Lewis 1992). Figure 3 shows the plot of two variates out of a synthetic 10-variate normal sample with very high uniform covariances $\sigma_{ij} = 0.99$. This case mimics the covariance structure between forecast series of adjacent predictor cells. The

sample was censored and successfully reconstructed by imputation.

Poor mixing is known to cause oscillation of the sampling during imputation, as visible from the trace plots in Fig. 4, and eventually to lead to the divergence of the iterative process. This difficulty can be overcome by diagonalizing the variance-covariance matrix $\boldsymbol{\Sigma}$ through principal component analysis (PCA).

PCA is a linear transformation and presupposes that the sufficient statistics mean and variance fully describe the sample distribution, a condition which is strictly met by Gaussian data and, as in our case, is obtained by normalizing precipitation data through NQT. We evaluate eigenvectors and obtain a diagonal covariance matrix with σ_{ii} given by eigenvalues. Eigen values represent the variance along principal components.

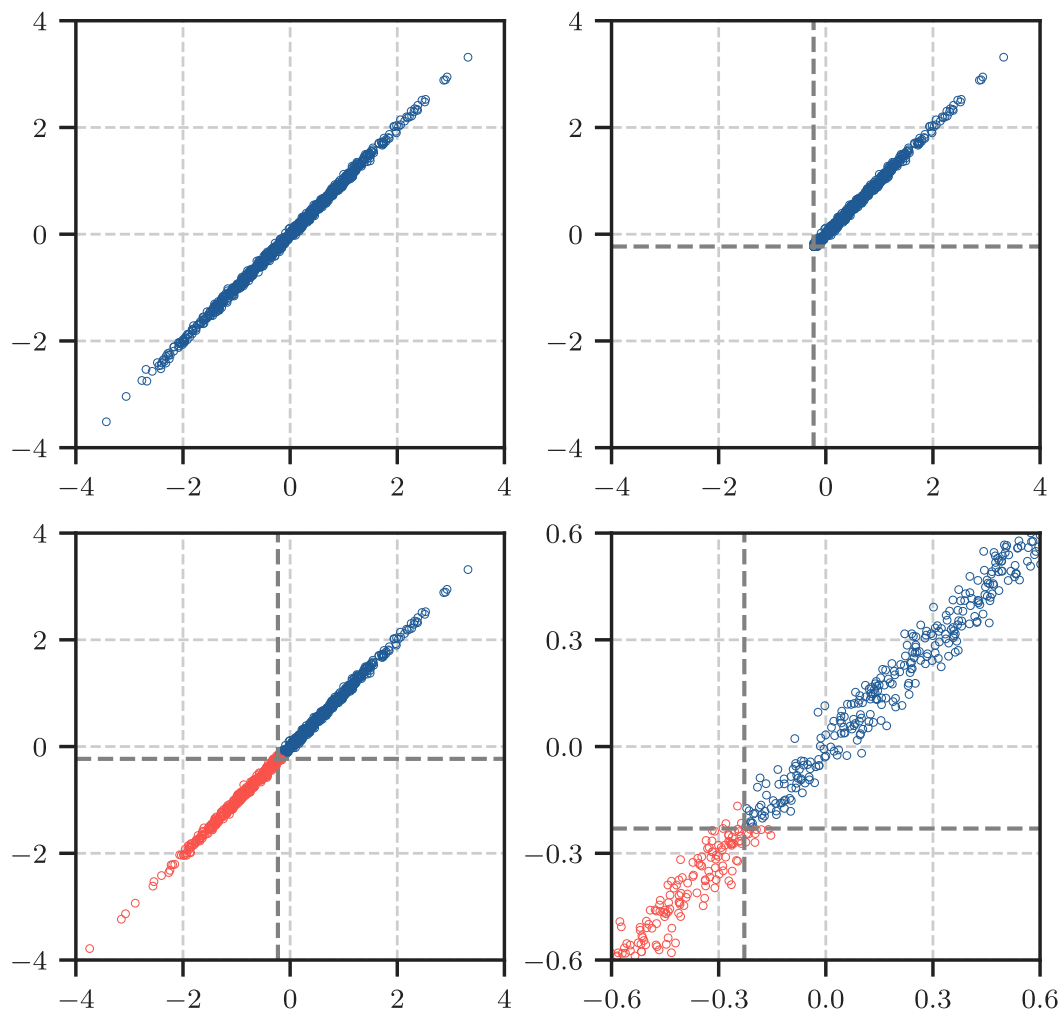


FIG. 3. Bayesian imputation for a highly correlated synthetic 10-variate normal sample, 10^3 draws. (top left) Variate 1 vs 2 of the original sample drawn from $N_{10}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = 0$, $\sigma_{ij} = 0.99$. (top right) Distribution truncated at the Gaussian values estimated from the Swiss data sample. (bottom left) Truncated sample with the red part retrieved by imputation, and (bottom right) a zoom in on the transition zone between parent and imputed sample. High correlation leads to poor mixing in the Gibbs sampling.

Principal components with larger associated variances indicate high informative content, while those with lower variances resemble noise. PCA allows ranking predictors in terms of their informative contribution and, as a consequence, reducing the dimensionality of the problem, a property which becomes of significant interest when working with ensemble forecasts.

An additional complication in this process is given by the fact that by performing PCA, one needs to consider the transformation by rotation of the entire sampling sector while mapping the 10-dimensional reference system into principal components. Figure 5 gives an example of this situation. We see a snapshot of sampling (Kotecha and Djurić 1999) from a bivariate truncated Gaussian distribution in the inner Gibbs

sampling cycle (appendix C). With reference to the composite variate \mathbf{Y} in the expression in (11), the selected snapshot arises from sampling at a given time step j , for which $r = 2$ elements $y_{j,k}^c$ indicate zero precipitation, while the remaining $m + 1 - r$ ones are nonzero. On a different time step one could have $r > 2$ elements equal to zero and the remaining ones nonzero, requiring to sample from an ($r > 2$)-dimensional truncated region. The employment of PCA involves the rotation of the sampling region, which leads to sampling constrained by a system of linear inequalities as explained by Li and Ghosh (2015). Last, but not least, PCA reduces the need for large numbers of burn-in steps of the Gibbs sampler, leading to considerable computational gain. Thanks to PCA the burn-in

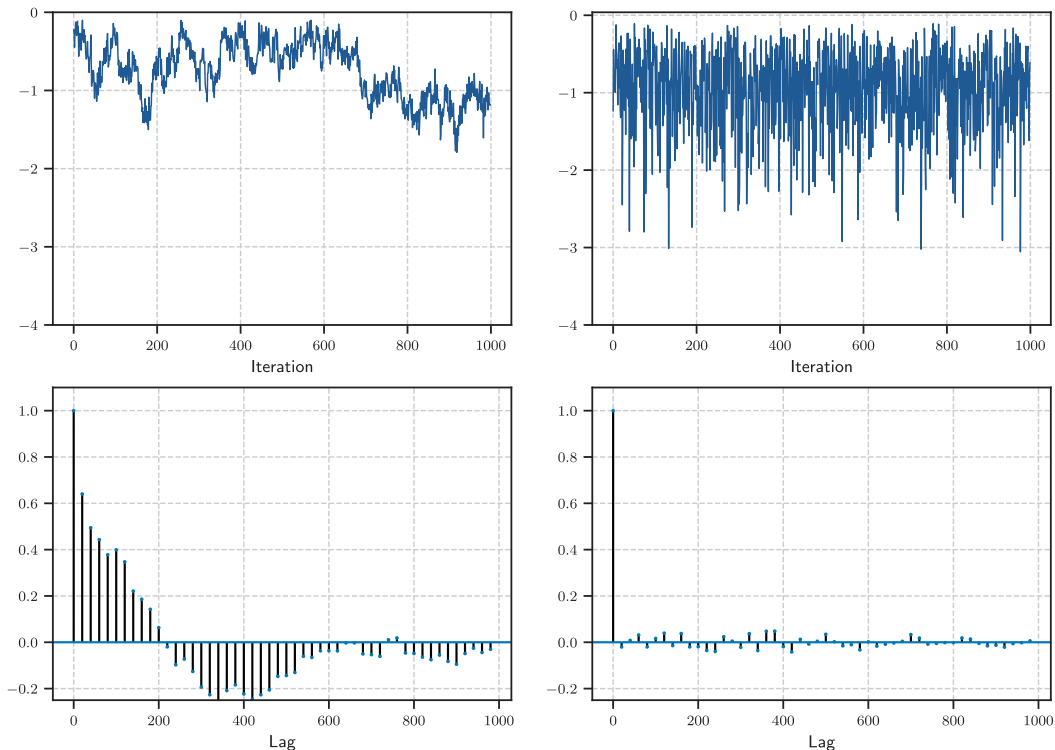


FIG. 4. Effects of poor mixing on Gibbs sampling for one variate of a 10-variate synthetic Gaussian sample, $\sigma_{ij} = 0.99$, 10^3 draws. (top) (left) the oscillating trace (a Gaussian variate) of the Gibbs sampler without PCA, and (right) the stabilization of the sampling due to PCA transformation, which diagonalizes the variance-covariance matrix ($\sigma_{ij} = 0.99$; $i \neq j$). (bottom) Autocorrelation function (ACF) of successive draws. Without PCA (left) undesired autocorrelation persists across the iterative sampling, while it decays after few iterations when adopting PCA.

steps in our case could be reduced from several thousands to few hundreds. After completing an outer Gibbs sampling iteration, the PCA transformation is inverted to retrieve the nondiagonal covariance matrix Σ .

e. Testing

Next we proceed to testing the validity of assumption in (6). Choosing to model the dependence structure of the joint normal sample (\mathbf{w}, \mathbf{z}) as MVN, is equivalent to using a Gaussian copula that is constructed starting from the corresponding Gaussian marginal distributions F and G . The MVN dependence is one of several possible dependencies we could have chosen, and of course is only an approximation of the true dependence structure of the joint sample, which we nevertheless assume sufficiently close to MVND. Therefore a positive outcome from stringent multivariate normality tests, such as the Mardia test (Mardia 1970), is illusory in our case, not so much because of the distance between the Gaussian copula and the MVND, but the presence of outliers, especially in the fringe region of the MVND associated with high or low-end extreme events.

Nevertheless, other tests can be performed to investigate sufficient closeness of the copula to a MVND. So the sample tested positively for the pairwise linear predictor versus predictor and predictand versus predictor dependence, as shown in the first row of Fig. 6 for an observation–predictor pair (left) and a predictor–predictor pair (right). We also tested the dependence of the residuals on the regressor using the Breusch–Pagan test (Breusch and Pagan 1979), which led to the acceptance of the homoscedasticity hypothesis for the majority of cases (second row in Fig. 6).

A Shapiro–Wilks test for normality of the residuals was performed, but led to rejection of some cases, despite residuals were visually on the theoretical Gaussian CDF curve. Especially the tails region of the pairwise Q–Q plots showed divergence from the bisection line (third row in Fig. 6).

Finally, we tested the Gibbs sampler on the 10-variate synthetic sample, which was drawn from the genuine MVND and, as to be expected, passed the Mardia test. After censoring the sample and retrieving the censored part by imputation, we performed another positive test for multivariate normality of the complete joint sample,

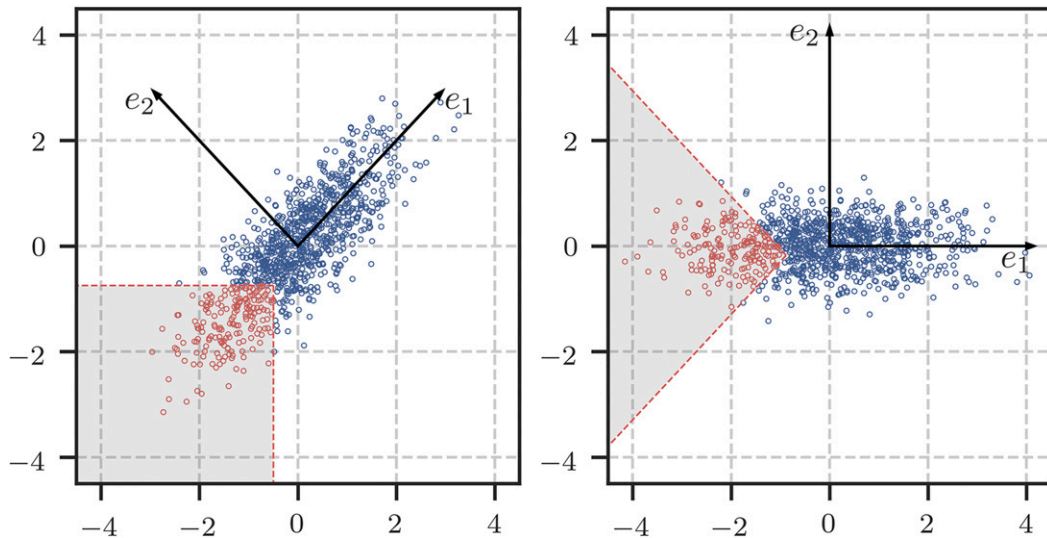


FIG. 5. Slice of the inner Gibbs sampling cycle for the 2D-case in which missing values are sampled from the bivariate truncated normal distribution. The reference frame is rotated by PCA transformation with eigenvectors on display. The red cloud is an ensemble of sampled fictitious negative precipitation values from which only one is retained at each inner cycling step. Due to rotation of the reference frame, the straight boundaries of the gray sampling region change accordingly. Thusly, instead of drawing from the (left) gray square region, one must draw from the (right) gray triangle region, which requires linear inequality constraining.

obtaining confirmation that the Gibbs sampler produces perfectly multivariate Gaussian data.

4. Results

a. Processor execution

The 10×13514 points study dataset, including 9 forecast and 1 observation series, is split into a 1 January 1979–31 December 2010 calibration and a 1 January 2011–31 December 2015 validation period. The processor is set up for the first period and verified over the second, whereby the conditional mean is compared against observations. This corresponds to a practical situation, in which a forecasting service uses a processor for the 5-yr period 2011–15 after it has been calibrated on 31 years of daily data and updated last on the 31 December 2010. The observed precipitation climatology is Weibull distributed, as visible in Fig. 7. The normalized subzero precipitation data, considered as censored negative precipitation, are retrieved by imputation for the calibration period. The sampling yields the posterior variance-covariance matrix Σ and a zero mean vector μ , to be used for processing of the validation period. First we apply (9) to compute conditional mean and variance, which is constant in virtue of near-homoscedasticity of the data. The Gaussian predictive density is computed by means of (8) and successively mapped back into the space of provenience. The

continuous Gaussian process must be transformed into a discontinuous process of type (A1). To this end the dichotomous PoP process is drawn from the Bernoulli distribution, a special case of the binomial distribution, while the precipitation depth process is drawn from the inverse Weibull.

The results of the back-transformation for a selected period are visualized in Fig. 8. Figure 8a depicts Gaussian data, including daily observed precipitation, the mean of the predictive distribution conditioned on a nonuple of predictors, and the 50%–90% credible intervals. The data below the Gaussian zero-precipitation threshold value of 0.06 (horizontal dashed black line) have been added by data augmentation. The green PDFs are the predictive densities for six selected days. Figure 8b compares the same data in the real space after back transformation. The PoP is sampled from the Bernoulli distribution by taking the area portion under the Gaussian PDF below threshold as distribution parameter $p \in [0, 1]$:

$$f(k|p) = p^k(1-p)^{1-k}; \quad k \in \{0, 1\}. \quad (28)$$

This area p represents the probability of binary precipitation occurrence/nonoccurrence, which is being randomized through the sampling process. Out of a r -sized sample of Bernoulli process realizations for a given day, drawing returns $q < r$ values of the event occurrence indicator equal to 1. For those q events,

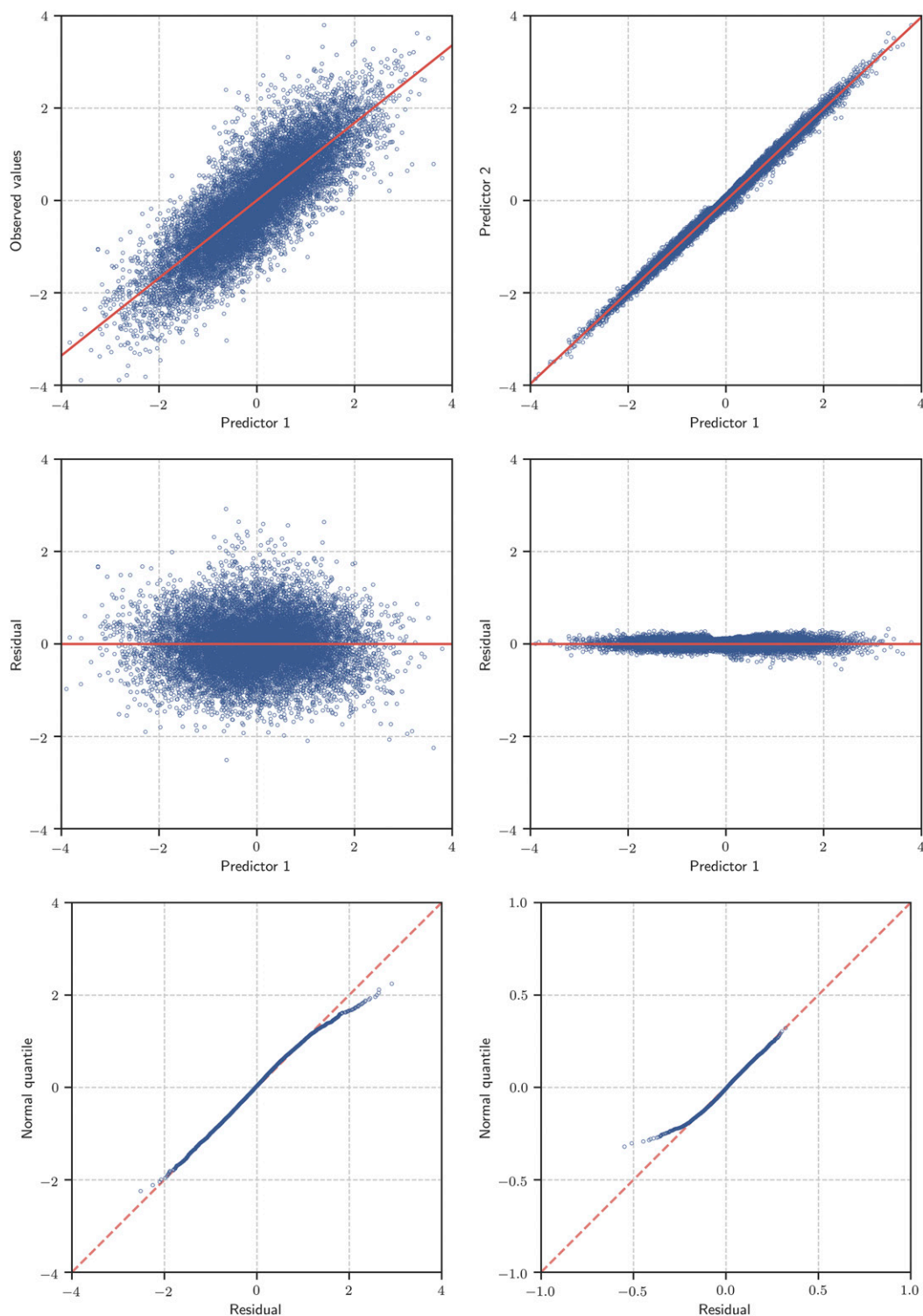


FIG. 6. Test results of multivariate normality for (left) an observations-predictor and (right) a predictor-predictor pair for selected predictors. (top) Linear interdependence, (middle) the residuals against a selected predictor, and (bottom) the Q-Q plot of the residuals. While linearity of the dependence and homoscedasticity of the residuals are visible, there is a divergence of the residuals from the bisection line in the tails region.

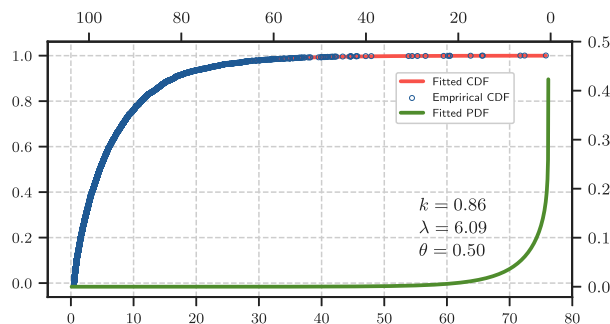


FIG. 7. Empirical and modeled climatic distribution of daily precipitation observed at BIZ, 1979–2015. The lower abscissa and left ordinate axis refer to the CDF, the other axes to the PDF. The data follow a 3-parameter Weibull distribution with k the shape, λ the scale, and θ the location parameter.

precipitation depth is drawn from the inverse Weibull distribution, while for the remaining $r - q$ events precipitation depth is set equal to zero. The predictive mean precipitation depth (red line in Fig. 8b) is

calculated as the average of the total r -sized sample of zero and non-zero-depth events. We note that the latter differs from the Gaussian conditional mean and median (the red line connecting 50% quantile points in Fig. 8a), both equal due to symmetry of the Gaussian distribution. From the sampled predictive distribution in the real space computation of credible intervals and sample variance are straight forward.

b. Verification metrics

In Table 1 we compare performance indicators for the uncertainty processor involving 9 predictors, one for each cell in Fig. 1, and the same processor involving only a single predictor at the central analysis cell. In this way we demonstrate the added value of involving multiple predictors to account for the spatial uncertainty of precipitation. As benchmark case for performance comparisons we use Bayesian model averaging (BMA) (Raftery et al. 2005). We applied BMA in the exact same way as our proposed Bayesian approach, using the same

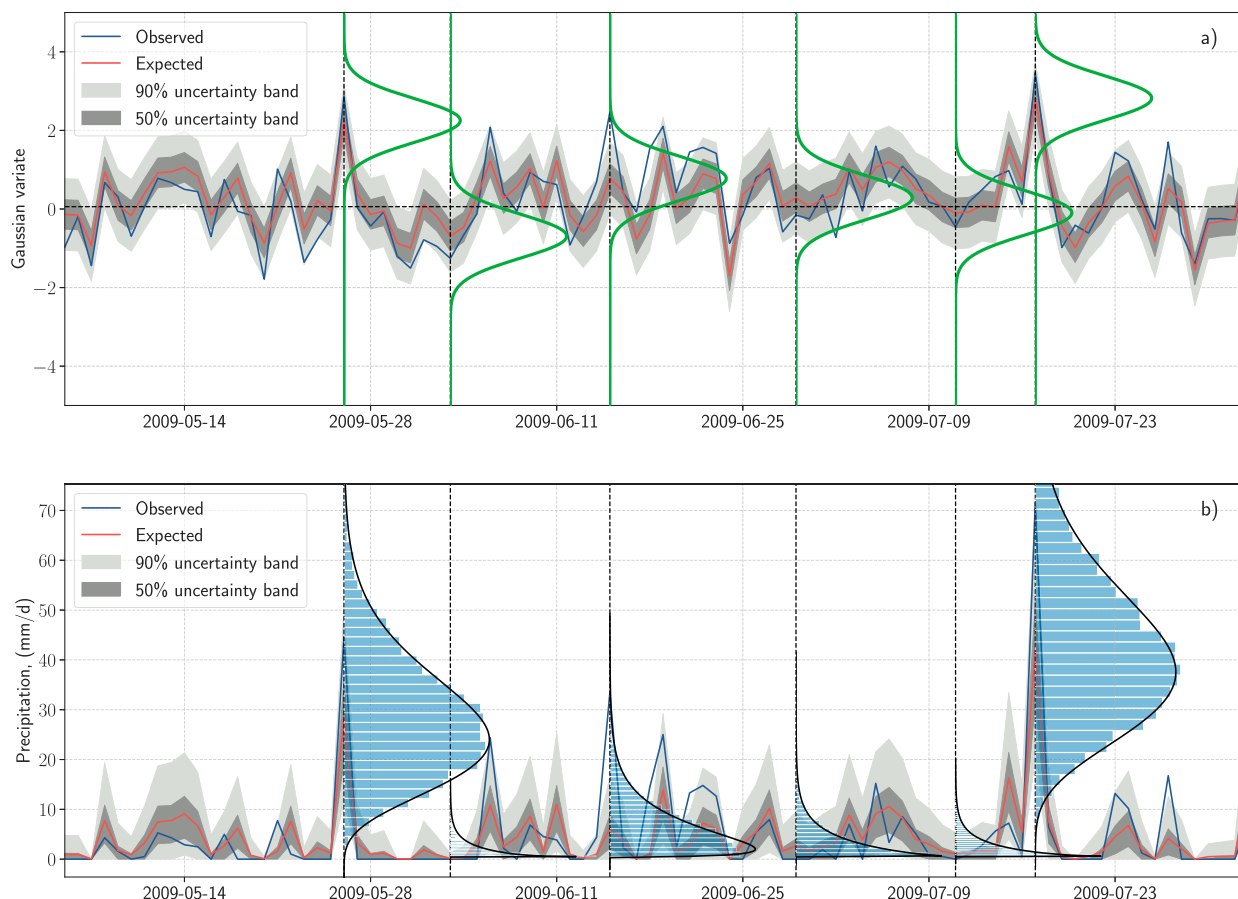


FIG. 8. (a) The observed precipitation, the conditional mean, and credible intervals for the Gaussian variables over a selected 4-month period at the central analysis cell BIZ (Fig. 1). The dashed horizontal line indicates the Gaussian zero-precipitation threshold. (b) The same data backtransformed into the original space. We note that the Gaussian PDFs morph into skewed gamma-type PDFs. The four Gaussian PDFs in the middle have part of the curve below the Gaussian zero-precipitation threshold. The area portions below the threshold are the PoP nonoccurrence used to parameterize the Bernoulli sampler for drawing the real-space PoP.

TABLE 1. Performance indicators, 9 cells vs 1 cell, 1979–2010 [calibration (cal)] vs 2011–15 [validation (val)] for proposed approach, BMA, and the raw unprocessed forecast at the central cell.

Method		Proposed approach				BMA		Raw	
Indicator	Definition	Cal	Val	Cal	Val	Cal	Val	Cal	Val
No. of cells		9	1	9	1	9	9	1	1
CORR(\mathbf{w} , $\boldsymbol{\mu}_{\mathbf{w} \mathbf{z}}$)	$E\{[\mathbf{w} - E(\mathbf{w})][\boldsymbol{\mu}_{\mathbf{w} \mathbf{z}} - E(\boldsymbol{\mu}_{\mathbf{w} \mathbf{z}})]\}$	0.84	0.83	—	—	0.83	—	—	—
VAR explained	$R^2 = 1 - \sigma_{\mathbf{w} \mathbf{z}}^2 = \boldsymbol{\Sigma}_{\mathbf{wz}} \boldsymbol{\Sigma}_{\mathbf{zz}}^{-1} \boldsymbol{\Sigma}_{\mathbf{zw}}$	0.71	0.69	—	—	0.68	—	—	—
VAR unexplained	$1 - R^2 = \sigma_{\mathbf{w} \mathbf{z}}^2 = 1 - \boldsymbol{\Sigma}_{\mathbf{wz}} \boldsymbol{\Sigma}_{\mathbf{zz}}^{-1} \boldsymbol{\Sigma}_{\mathbf{zw}}$	0.29	0.31	—	—	0.32	—	—	—
SNR	$R^2/(1 - R^2)$	2.40	2.19	—	—	2.16	—	—	—
BIAS	$(1/n) \sum_{i=1}^n (x_i - \mu_{\mathbf{x} \mathbf{x}_i})$	0.13	0.31	0.32	0.26	0.13	0.26	−0.39	−0.2
MAE	$(1/n) \sum_{i=1}^n x_i - \mu_{\mathbf{x} \mathbf{x}_i} $	2.26	2.34	2.24	2.32	2.35	2.33	2.44	2.37
RMSE	$\left[(1/n) \sum_{i=1}^n (x_i - \mu_{\mathbf{x} \mathbf{x}_i})^2 \right]^{1/2}$	4.45	4.50	4.55	4.57	4.50	4.57	4.72	4.7
CORR(\mathbf{x} , $\boldsymbol{\mu}_{\mathbf{x} \mathbf{x}}$)	$E\{[\mathbf{x} - E(\mathbf{x})][\boldsymbol{\mu}_{\mathbf{x} \mathbf{x}} - E(\boldsymbol{\mu}_{\mathbf{x} \mathbf{x}})]\}$	0.72	0.71	0.71	0.71	0.71	0.71	0.71	0.71
CRPS	$(1/n) \sum_{i=1}^n \text{CRPS}(F_i, x_i)$	1.54	1.60	1.55	1.59	1.6	1.58	—	—

NQT-normalized variates, which were extended into the fictitious subzero precipitation range by imputation. Then Gaussian univariate normal densities of precipitation, conditional on individual forecasts for the 9 predictor cells are obtained by exploiting the properties of bivariate normal distributions (Mardia et al. 1979), and the conditional densities mixed by linear weighting:

$$\mathcal{L}(\gamma_1, \dots, \gamma_n | \mathbf{w}, \mathbf{z}_1, \dots, \mathbf{z}_n) = \sum_{i=1}^n \log \left[\sum_{j=1}^m \gamma_j \phi(w_i | z_{ij}) \right]$$

subject to: $\sum_{j=1}^m \gamma_j = 1,$ (29)

where n is the number of temporal steps and m is the number of predictors. The weights γ_i are calculated by maximizing the log-likelihood function \mathcal{L} through expectation maximization (EM) or a Newton–Raphson solver. BMA estimates mean and variance of the predictive density through weighted combination of mean and variance of the constituent predictive densities, once weights are known. As BMA application for a single cell is not meaningful, we calculate the corresponding performance indicators for the 9-cell case only.

Table 1 is split into two parts, an upper one with indicators calculated in the Gaussian space and a lower one with indicators in the real space. Vertically the table compares results for the proposed method, BMA and the raw unprocessed prediction at the central analysis cell. For the proposed method the Pearson correlation (CORR) between observations and conditional mean, which is coincident with the covariance for standard Gaussian variates, computes at 0.84 for 9 predictors, indicating good agreement between Gaussian observations and the linear model. The value is slightly smaller

for the application with a single predictor. Next, we report the coefficient of multiple correlation R^2 (coefficient of determination), which can be interpreted as the variance (VAR) of the observations estimated solely from the regression model, and the variance of residuals in (9) (variance unexplained), equal to $1 - R^2$. The values corroborate that the linear regression model in (4) is able to explain 71% of the Gaussian variance, while 29% remains random noise. We do not report these values for the verification period, as the processor continues to operate with the parameters ($\boldsymbol{\mu}_{\mathbf{w}|\mathbf{z}}$, $\boldsymbol{\Sigma}_{\mathbf{zz}}$) retrieved for the calibration period. It is possible to compute these parameters retrospectively for the validation period, but this would require another round of imputation to recover subzero observations and predictions, while the difference with respect to the calibration period would be likely insignificant. The results for BMA are very similar, but slightly worse. The BMA variance averaged over the calibration and the validation period computes to 0.32, thus slightly larger than the variance given by our approach. The signal-to-noise ratio (SNR) is a decision-theoretic measure of the informativeness of output (Krzyzstofowicz 1992) and is equal to 2.4 for the proposed method and 2.16 for BMA due to the slightly higher variance. In the hypothetical case of a totally uninformative forecast, which would be completely uncorrelated with observations, $\text{CORR}(\mathbf{w}, \boldsymbol{\mu}_{\mathbf{w}|\mathbf{z}}) \approx 0$, all variance becomes unexplained and $\text{SNR} \rightarrow 0$. To the contrary, if the processed forecast is “perfect,” $\text{CORR}(\mathbf{w}, \boldsymbol{\mu}_{\mathbf{w}|\mathbf{z}}) = 1$, and consequently $\text{SNR} \rightarrow \infty$. This also means that in the case of a non-informative forecasting model, which poorly correlates with observations, $\boldsymbol{\Sigma}_{\mathbf{wz}} \boldsymbol{\Sigma}_{\mathbf{zz}}^{-1} \boldsymbol{\Sigma}_{\mathbf{zw}} \rightarrow 0$. The conditional mean collapses onto the climatological mean and the variance approaches that of retrospective observations,

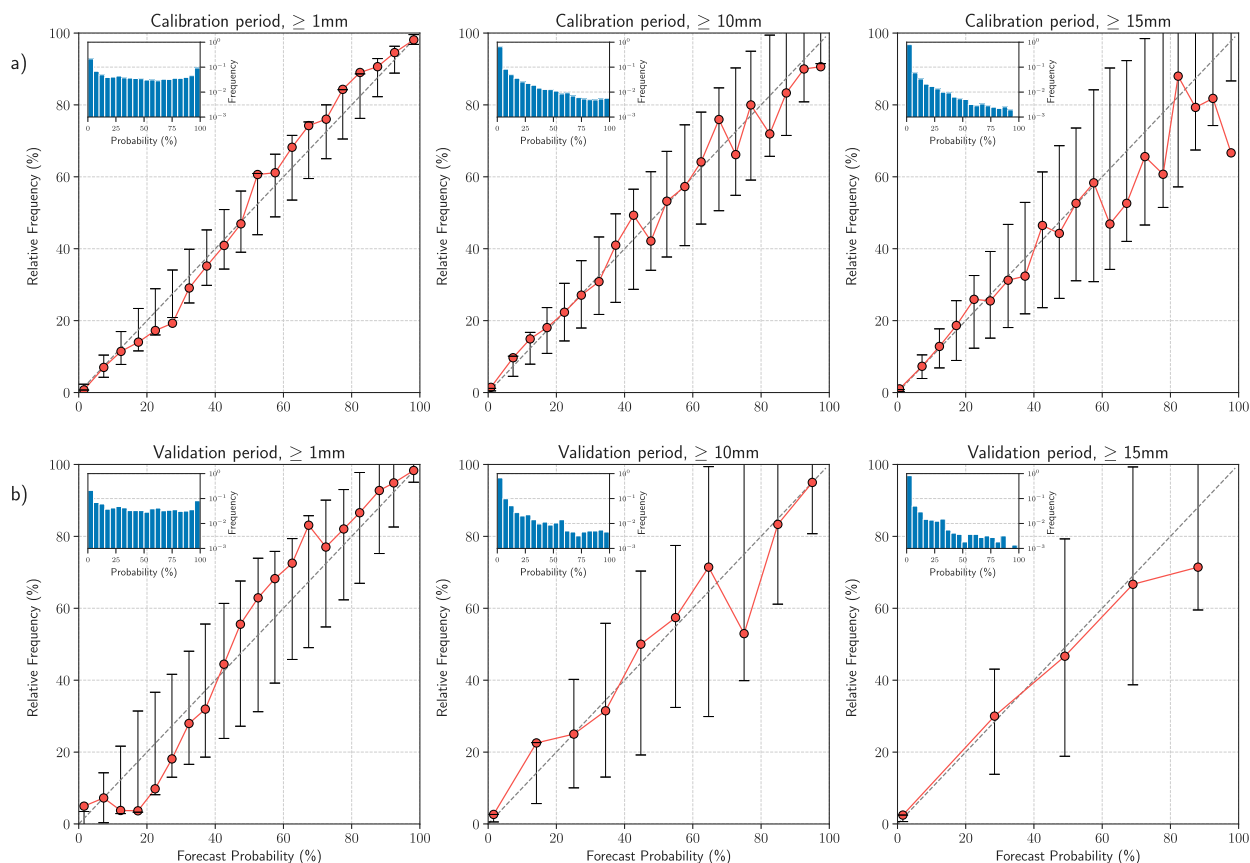


FIG. 9. Reliability diagrams for (a) calibration and (b) verification, threshold values $V = [1, 10, 15]$ mm day $^{-1}$, 9 predictors. The insets visualize relative frequencies of the forecast distribution $p(q_i)$. With increasing threshold V , calibration deteriorates visibly through departure from the bisection in either direction. In the first panel in (b) the marker at 0% on the ordinate axis means that particular bin contains only observation nonoccurrences. Because of rarefaction of observed frequencies in the high precipitation range for validation, the number of bins has been reduced from 20 to 5. The vertical bars indicate consistency intervals. Observed relative frequencies are in nearly all cases within the 5%–95% bounds and thus consistent with reliability.

precluding the production of a probabilistic forecast which is less informative than climatology and yields negative economic value. This internal coherence property is an essential requirement for using forecasts in rational decision making (Krzysztofowicz 1999).

The lower part of the table reports bias (BIAS), mean absolute error (MAE), root-mean-square error (RMSE) and correlation (CORR), metrics that are all estimated in the real variable space. We note that one effect of the Bayesian processor is bias removal, which is reduced well below 0.5 mm day $^{-1}$ in all cases.

Processor performance is distilled through reliability diagrams (Wilks 1995; Bröcker and Smith 2007), which visualize dicotomic occurrence/nonoccurrence frequencies of an observation against the probability of the corresponding forecast. First we preset precipitation thresholds $V = [1, 10, 15]$ mm day $^{-1}$ against which dicotomic occurrence/nonoccurrence is verified. Given V , a forecast is considered reliable if actual

precipitation exceeding V occurs with an observed relative frequency consistent with the forecast value. If o_j are occurrence/non-occurrence frequencies of daily observations and q_i the corresponding allowable probabilities of forecasts exceeding V , the reliability diagram collapses the joint distribution $p(o_j, q_i)$ by factorization into the conditional distribution $p(o_j|q_i)$ (calibration distribution) and the forecast distribution $p(q_i)$ (refinement distribution) (Murphy and Winkler 1987). The relative frequency of observations are plotted against suitably binned forecast probability quantiles. The forecast distribution is visualized as inset frequency histogram on the same plot. Figure 9 shows the reliability diagrams for validation (Fig. 9a) and calibration (Fig. 9b) periods for different values V . An ideally calibrated processor producing perfect forecasts leads to a graph with markers lying on the bisection line. Markers below the bisection indicate systematic overforecasting and thus wet bias, while those above

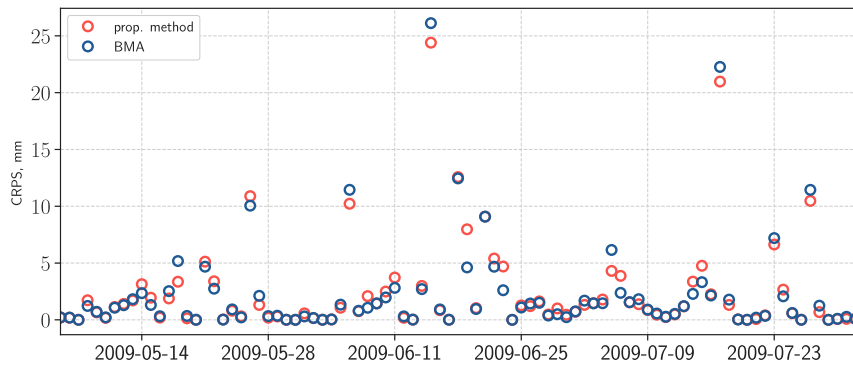


FIG. 10. Daily CRPS values for the same selected 4-month period in Fig. 8, red for the proposed method, blue for BMA. The $\overline{\text{CRPS}}$ for this period is 2.40 and 2.36, respectively, for the two methods.

mean underforecasting and dry bias. A forecast that is biased in either of the two directions is considered unreliable or miscalibrated.

The vertical credible interval bars in Fig. 9 are derived as in Bröcker and Smith (2007) by computing variations of the observed relative frequencies over a set of forecasts generated by bootstrap resampling with replacement. The method computes variations of the observed relative frequencies resulting from uncertainties in terms of quantile bin mean probability and bin population. The result of such sampling produces a frequency distribution in correspondence of each bin, from which the 5%–95% credible interval is determined. The bin size in Fig. 9b has been enlarged in the high precipitation range $>10 \text{ mm day}^{-1}$ to avoid meaningless observed relative frequencies due to entirely full or empty quantile bins.

The drawback of such condensed representation is the inability to provide a full description of the forecast quality. For a more fine-grained event-based forecast performance investigation we evaluate the continuous ranked probability score (CRPS) (Matheson and Winkler 1976). The CRPS is the integral of the Brier score over all possible threshold values v for a continuous predictand, given a forecast realization. Specifically, if F is the predictive CDF and x the benchmark observation, the continuous ranked probability score is defined as follows:

$$\text{CRPS}(F, x) = \int_{v=0}^{v=\infty} [F(\xi) - 1(\xi \geq x)]^2 d\xi, \quad (30)$$

where $1(\xi \geq x)$ denotes the Heaviside function with value 0 when $\xi < x$ and 1 otherwise. Equation (30) represents an integral distance measure between the predictive CDF and the Heaviside function. In our case the cumulative distribution function F is not available in closed form and must be evaluated from the predictive

density, which is given by an ensemble of discrete points making up the frequency histograms in Fig. 8b. The average CRPS over n discrete events is calculated as follows:

$$\overline{\text{CRPS}} = \frac{1}{n} \sum_{i=1}^n \text{CRPS}(F_i, x_i). \quad (31)$$

For an intuitive understanding of the meaning of (31), Hersbach (2000) demonstrated that in the particular case of single deterministic forecast $F_i = F(x_i | \hat{x}_{\text{det},i})$ degenerates into a step function and as a result $\overline{\text{CRPS}}$ becomes the MAE, which has a clear interpretation. In the probabilistic case the CRPS represents a generalization of the MAE (Gneiting et al. 2005). Figure 10 shows daily $\text{CRPS}(F_i, x_i)$ values for a selected period of four consecutive months for the proposed method and for BMA. The CRPS computes to 2.40 and 2.36, respectively. Table 1 reports the $\overline{\text{CRPS}}$ for the calibration and the validation period given 1 and 9 predictors for the proposed method and 9 predictors for BMA. Values are very close. We also compared the CDFs of the CRPS for the proposed method against BMA, calibration and validation period. The CDFs are near-coincident, making a display superfluous.

5. Discussion

Thus far we presented the methodology and application of the proposed Bayesian uncertainty processor, which is founded on the model conditional processor (MCP) concept (Todini 2008). In principle there is neither theoretical nor procedural limitation to the number of employable predictors thanks to working with normal variables and a nonparametric structure. Additional predictors can easily be included into the analysis.

The proposed Bayesian processor has been benchmarked against BMA and the performance indicators in

Table 1 figure in the same range, with the BMA average variance over the calibration and the validation period slightly higher. Unlike Slougher et al. (2007), who derived heuristic conditional densities of precipitation ad hoc as logistic regressions with a power transformation, we have first normalized variates by nonparametric NQT, performed data augmentation and successively applied BMA to Gaussian data. Todini (2008) and Biondi and Todini (2018) pursued a similar performance contest of the two approaches, albeit in a different context, confirming strong similarities between BMA and the proposed Bayesian approach in terms of predictive mean and a slightly higher variance for BMA. We also note that BMA constitutes an approximation of predictive mean and variance, requiring constrained optimization for determining the weights. Moreover, BMA does not explicitly consider the covariance structure among predictors. Our proposed method instead estimates predictive distributions on an analytical basis by nonparametric mapping of empirical distributions to the Gaussian space, while accounting for the dependency structure among predictors. A certain level of approximation nevertheless remains due to assuming a MVN dependency among NQT-transformed variates and estimating the covariance structure by MCMC sampling.

Another topic of discussion is our choice of the spatial set of predictors. These are given by grid-based precipitation forecasts in a 9-cell modeling window centered on an analysis cell. This choice is motivated by the necessity to account for spatial uncertainty of precipitation. The methodology does not limit the extension of the analysis mask to larger windows as the one chosen here. Moreover, the approach can be spatialized by applying a sliding analysis mask over larger regions. Such spatial use of the processor requires estimating the covariance matrix Σ_{zz} and the covariance vector Σ_{wz} on a cell-by-cell basis for the study region. Of course these quantities need to be re-estimated periodically as new observations and forecasts become available. At an operational level such updating, which necessitates Bayesian imputation, can be executed offline, without impacting online operations.

We also note that the processor has been calibrated over a single dataset, without slicing it by seasons or specific pluvial regimes. The MVND should correctly account for extreme events and the multivariate normal regression model thus fitted separately for a low-to-intermediate and a high precipitation range. Such an approach would improve the estimation of extreme precipitation events, which tend to be underrepresented in the current setup. Examples of calibrating the processor by means of multivariate truncated normal

distributions to accommodate heteroscedasticity of the data are given in Coccia and Todini (2011) and Reggiani et al. (2016, 2019).

An aspect, which we believe deserves some further consideration is the application of principal component analysis (PCA) that was used to diagonalize Σ_{zz} and optimizing the MCMC sampling. In our study data example the 10×10 dimensional variance-covariance matrix reduces to the following ranked eigenvalue vector:

$$\text{diag}(9.94, 0.3, 0.02, 5.8 \times 10^{-3}, 1.0 \times 10^{-4}, 1.0 \times 10^{-4}, 0.0, \dots) \quad (32)$$

indicating that the problem can be considerably downsized, as two principal components explain most of the variance. While we recognize that in our study case the problem is strongly reducible, there may be different locations, for which a larger number of dimensions need to be retained. Examples are locations, in which precipitation is linked to orographic effects or particular predominant air currents. PCA, which must be performed on strictly Gaussian variates, is a powerful approach to objectively identify the minimum number of dimensions required for forecast postprocessing. Such an approach becomes particularly appealing when extending the processor to ensemble forecasting.

6. Summary and conclusions

The outlined methodology describes a precipitation uncertainty processor able to handle multiple binary-continuous predictors. The output of the processor is a calibrated binary-continuous debiased probabilistic forecast of precipitation at a single location. Precipitation, an intermittent random process, is considered as censored variate, turned into a continuous one by recovering unknown censored values by imputation. Processing of the predictors and the observed values starts with the transformation of nonparametric marginal distributions into standard normal variates by NQT. The normalization serves different purposes.

First, the joint distribution can be considered with some approximation as MVND, equivalent to a Gaussian copula, which admits closed-form expressions for conditional densities.

Second, the censored Gaussian precipitation values and MVND parameters are retrieved by Bayesian imputation using a nested Gibbs sampler. The use of fully Gaussian distributed data facilitates the monitoring of imputation convergence and verification of results.

Third, Gaussianity of the data supports the application of principal component analysis (PCA) to

diagonalize the variance-covariance matrix and analyze data redundancy.

Conditional predictive densities are computed for the normal variables and successively back-transformed into the space of origin. The processor has been calibrated and validated for a test site in Switzerland and computes satisfying reliability plots as well as performance indicators against BMA. The principal strengths of the proposed method can be summarized as follows:

- The processor does not use ad hoc assumptions on distribution models for precipitation depth, but works straight with empirical CDFs that are subsequently mapped to Gaussian by nonparametric transformations. This supports parameter parsimony and avoids the need for parameter optimization.
- Being parameter-parsimonious, the processor is computationally efficient and consequently apt for operational use.
- The processor is sufficiently generic to handle multiple predictors, encouraging its application for hydrometeorological applications involving similar intermittent random processes, for example river flows.
- The processor is self-calibrating given it produces an output (the predictive mean) with the same distributional properties as retrospective observations.
- Moreover, the processor guarantees coherence as it cannot produce an output with inferior informativeness and thus negative economic value than the usage of the climatic distribution.
- Standard forecast verification shows that the processor meets quality criteria, such as bias removal, and yields comparable values against BMA in terms of commonly used performance metrics.

Acknowledgments. This research was supported by Deutsche Forschungsgemeinschaft through Grant RE3834/5 “BSCALE” awarded to the first author. We thank Ezio Todini for the BMA likelihood maximization algorithm and Alfred Müller from the University of Siegen for constructive suggestions. Both have supported this work with their discussions. We also acknowledge Meteo Swiss and ECMWF for giving access to the data used in this study. We finally acknowledge three anonymous reviewers for their thorough reading and suggestions that have helped improving the manuscript.

APPENDIX A

Precipitation Process Representation

Stochastically the precipitation process is described as mixture of binary and continuous variates. If the variate

\mathbf{X} describes precipitation depth accumulation with realization \mathbf{x} , $(1 - \nu) = P(\mathbf{X} > 0)$ is the dichotomous probability of precipitation (PoP), while $H_o(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x} | \mathbf{X} > 0)$ is the probability of the continuous precipitation depth accumulation process, such that $H_o \geq 0$ if $\mathbf{x} > 0$ and $H_o = 0$ if $\mathbf{x} = 0$, the combined probability $H(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x} | \mathbf{X} \geq 0)$ of the mixed binary-continuous process is (Kelly and Krzysztofowicz 2000):

$$H(\mathbf{x}) = \nu + (1 - \nu) H_o(\mathbf{x}), \quad (\text{A1})$$

which assigns a probability mass of ν to event $\mathbf{X} = 0$ and spreads the remaining mass $(1 - \nu)$ over the interval $(0, \infty)$. The continuous process is modeled with a suitable parametric cumulative distribution function (e.g., Weibull, Beta, or Gamma), while event occurrence can be estimated as climatological relative frequency from a sample of observations, or modeled as a stochastic process, for instance a Markov chain (Katz 1977; Woolhiser and Pegram 1979), a Poisson (Todorovic and Yevjevich 1969; Gupta and Duckstein 1975), a Neymann–Scott (Kavvas and Delleur 1981; Waymire and Gupta 1981) or a Pólya urn process (Todini and Di Bacco 1997).

If the precipitation is forecast by a single predictor, the mixture includes dependence on the predictor variate $\hat{\mathbf{X}}$ (Herr and Krzysztofowicz 2005):

$$H(\mathbf{x}, \hat{\mathbf{x}}) = \hat{\nu}_{00} + \hat{\nu}_{10} H_{\hat{\mathbf{x}}}(\mathbf{x} | 0) + \hat{\nu}_{01} H_{\hat{\mathbf{x}}}(\hat{\mathbf{x}} | 0) + \hat{\nu}_{11} H_{\mathbf{x}, \hat{\mathbf{x}}}(\mathbf{x}, \hat{\mathbf{x}}), \quad (\text{A2})$$

with $\hat{\nu}_{11} = P(\mathbf{X} > 0, \hat{\mathbf{X}} > 0)$, $\hat{\nu}_{01} = P(\mathbf{X} = 0, \hat{\mathbf{X}} > 0)$, $\hat{\nu}_{10} = P(\mathbf{X} > 0, \hat{\mathbf{X}} = 0)$, and $\hat{\nu}_{00} = P(\mathbf{X} = 0, \hat{\mathbf{X}} = 0)$ joint probabilities subject to $\hat{\nu}_{00} = 1 - \hat{\nu}_{01} - \hat{\nu}_{10} - \hat{\nu}_{11}$. The predictive density given by (1) and associated with (A2) is

$$h(\mathbf{x} | \hat{\mathbf{x}}) = \frac{\hat{\nu}_{11} h_{\mathbf{x}, \hat{\mathbf{x}}}(\mathbf{x}, \hat{\mathbf{x}})}{\hat{\nu}_{01} h_{\hat{\mathbf{x}}}(\hat{\mathbf{x}} | 0) + \hat{\nu}_{11} g_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})}, \quad (\text{A3})$$

where $h_{\mathbf{x}, \hat{\mathbf{x}}}(\mathbf{x}, \hat{\mathbf{x}})$ is the bivariate, $h_{\hat{\mathbf{x}}}(\hat{\mathbf{x}} | 0)$ is the conditional and $g_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})$ is the marginal density, and $h(\mathbf{x} | \bullet)$ is a family of truncated univariate conditional densities defined in $(0, \infty)$. In this paper we propose an alternative approach of determining $h(\mathbf{x} | \bullet)$, given a realization $\hat{\mathbf{X}} = \hat{\mathbf{x}}$ and generalizable for multiple predictors.

APPENDIX B

Normal Quantile Transform

Non-Gaussian random data, represented by variate \mathbf{X} with realization \mathbf{x} and CDF F , can always be mapped into a standard normal variate \mathbf{U} with realization \mathbf{u}

through a normal quantile transformation (NQT). The NQT is a strictly monotone nonparametric transformation T such that $T(\mathbf{X})$ is standard normal with CDF Φ and density ϕ . Thus the following relationship holds:

$$\begin{aligned}\Phi(\mathbf{u}) &= P[T(\mathbf{X}) \leq \mathbf{u}] = P[\mathbf{X} \leq T^{-1}(\mathbf{u})] \\ &= F[T^{-1}(\mathbf{u})] = F(\mathbf{x}).\end{aligned}\quad (\text{B1})$$

By applying the NQT to a censored or a truncated random variable \mathbf{X} one obtains a censored or truncated standard normal variate. Monotonicity of T ensures the existence of an inverse T^{-1} and thus a unique correspondence with the original variates.

APPENDIX C

Posterior Gibbs Sampling

Posterior distributions in (24) constitute normal conditional distributions, from which we can successively sample censored data and their mean and covariance parameters $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. At each sampling step we take the j th row vector $\mathbf{y} = [w_j, z_{j1}, \dots, z_{jm}]$ of the matrix $\mathbf{Y} = (y_{j,k}) = (\mathbf{w}, \mathbf{z})$ defined in (11). The length of the \mathbf{y} row subvector containing censored data, \mathbf{y}^c , is r , while the observed part \mathbf{y}^o has length $m + 1 - r$ [i.e., $\mathbf{y} = (\mathbf{y}^o, \mathbf{y}^c)$]. We start the “outer” MCMC sequence t for each of the n row vectors \mathbf{y} (indices j are omitted for notational simplicity) by initializing, given standard normality, four submatrices of the row-wise covariance matrices as identity matrices and the two means with zero. The covariance matrix between \mathbf{y}^o and \mathbf{y}^c is assumed zero at $t = 0$ (i.e., the two vectors are a priori uncorrelated):

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{y}^o, \mathbf{y}^o} &= \mathbf{I}, \quad \boldsymbol{\mu}_{\mathbf{y}^o} = 0 \\ \boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^c}^{(0)} &= \mathbf{I}, \quad \boldsymbol{\mu}_{\mathbf{y}^c}^{(0)} = 0 \\ \boldsymbol{\Sigma}_{\mathbf{y}^o, \mathbf{y}^c}^{(0)} &= \boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^o}^{(0)} \approx \mathbf{0}.\end{aligned}\quad (\text{C1})$$

The conditional mean vector and covariance matrix for $\mathbf{y}^{c,(t)}$ is calculated by making use of the properties of multivariate normal distributions (Mardia et al. 1979):

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{y}^c | \mathbf{y}^o}^{(t)} &= \boldsymbol{\mu}_{\mathbf{y}^c}^{(t)} + \boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^o}^{(t)} \boldsymbol{\Sigma}_{\mathbf{y}^o, \mathbf{y}^o}^{-1(t)} (\mathbf{y}^o - \boldsymbol{\mu}_{\mathbf{y}^o}^{(t)})^T \\ \boldsymbol{\Sigma}_{\mathbf{y}^c | \mathbf{y}^o}^{(t)} &= \boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^c}^{(t)} - \boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^o}^{(t)} \boldsymbol{\Sigma}_{\mathbf{y}^o, \mathbf{y}^o}^{-1(t)} \boldsymbol{\Sigma}_{\mathbf{y}^o, \mathbf{y}^c}^{(t)}.\end{aligned}\quad (\text{C2})$$

Now we sample $y_i^{c,(t)}$, $i = 1, \dots, r$ from the MVND right truncated at threshold value c (Li and Ghosh 2015):

$$y_i^{c,(t)} \sim N(\boldsymbol{\mu}_{\mathbf{y}^c | \mathbf{y}^o}^{(t)}, \boldsymbol{\Sigma}_{\mathbf{y}^c | \mathbf{y}^o}^{(t)}, -\infty, c); \quad i = 1, \dots, r. \quad (\text{C3})$$

This “inner” sampling is performed by means of a nested Gibbs sampler, as recovery of the censored observations implies simulation from the corresponding truncated multivariate normal distribution (Kotecha and Djurić 1999) by conditioning on y_i^o , $\boldsymbol{\mu}_{\mathbf{y}^c | \mathbf{y}^o}^{(t)}$, $\boldsymbol{\Sigma}_{\mathbf{y}^c | \mathbf{y}^o}^{(t)}$.

Next, a prior estimate of $\boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^c}^{(t)}$ is drawn from the inverse Wishart distribution (Gelman et al. 2014). The inverse Wishart distribution by definition always leads to real-valued positive-definite matrices. The latter is obtained by first estimating mean and sum of squares about the mean from the data and then using the estimate to draw the prior covariance matrix:

$$\begin{aligned}\bar{\mathbf{y}}^c &= E(\mathbf{y}^{c,(t)}) \\ \bar{\mathbf{S}}_{\mathbf{y}^c, \mathbf{y}^c} &= (\mathbf{y}^{c,(t)} - \bar{\mathbf{y}}^c)^T (\mathbf{y}^{c,(t)} - \bar{\mathbf{y}}^c) \\ \boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^c}^{(t)} &\sim \text{invWis}(r - 1, \bar{\mathbf{S}}_{\mathbf{y}^c, \mathbf{y}^c}^{-1}).\end{aligned}\quad (\text{C4})$$

Last we draw the mean $\boldsymbol{\mu}_{\mathbf{y}^c}^{(t)}$ from the multivariate normal distribution:

$$\boldsymbol{\mu}_{\mathbf{y}^c}^{(t)} \sim N(\bar{\mathbf{y}}^c, \boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^c}^{(t)}) \quad (\text{C5})$$

and calculate $\boldsymbol{\Sigma}_{\mathbf{y}^c, \mathbf{y}^o}^{(t)}$ from the sample. This operation is performed for all rows $j = 1, \dots, n$ of \mathbf{Y} . Then we return to (C2) and execute the subsequent sampling step. For all rows the first 500 “outer” Gibbs sampling steps are regarded as burn-in phase and are disregarded. For the “inner” nested Gibbs sampler 100 burn-in steps have been recognized as fully sufficient.

REFERENCES

- Alpert, M., and H. Raiffa, 1982: A progress report on the training of probability assessors. *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky, Eds., Cambridge University Press, 294–305, <https://doi.org/10.1017/CBO9780511809477.022>.
- Bárdossy, A., and E. Plate, 1992: Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res.*, **28**, 1247–1259, <https://doi.org/10.1029/91WR02589>.
- , and G. G. S. Pegram, 2009: Copula based multisite model for daily precipitation simulation. *Hydrol. Earth Syst. Sci.*, **13**, 2299–2314, <https://doi.org/10.5194/hess-13-2299-2009>.
- Berrisford, P., and Coauthors, 2011: The ERA-Interim archive version 2.0. Tech. Rep., European Centre for Medium-Range Weather Forecasts, 27 pp., <https://www.ecmwf.int/node/8174>.
- Biondi, D., and E. Todini, 2018: Comparing hydrological post-processors including ensemble predictions into full predictive probability distribution of stream flow. *Water Resour. Res.*, **54**, 9860–9882, <https://doi.org/10.1029/2017WR022432>.
- Breusch, T. S., and A. R. Pagan, 1979: A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294, <https://doi.org/10.2307/1911963>.
- Bröcker, J., and L. A. Smith, 2007: Increasing the reliability of reliability diagrams. *Mon. Wea. Rev.*, **22**, 651–662, <https://doi.org/10.1175/WAF993.1>.

- Coccia, G., and E. Todini, 2011: Recent developments in predictive uncertainty assessment based on the Model Conditional Processor approach. *Hydrol. Earth Syst. Sci.*, **15**, 3253–3274, <https://doi.org/10.5194/hess-15-3253-2011>.
- Cressie, N., 1985: Fitting variogram models by weighted least squares. *Math. Geol.*, **17** (5), 663–586.
- Frost, A. J., M. A. Thyer, R. Srikantan, and G. Kuczera, 2007: A general Bayesian framework for calibrating and evaluating stochastic models of annual multi-site hydrological data. *J. Hydrol.*, **340**, 129–148, <https://doi.org/10.1016/j.jhydrol.2007.03.023>.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. H. Vehtari, and D. B. Rubin, 2014: *Bayesian Data Analysis*. 3rd ed. CRC Press, 639 pp.
- Geman, S., and D. Geman, 1984: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-6**, 721–741, <https://doi.org/10.1109/TPAMI.1984.4767596>.
- Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, <https://doi.org/10.1175/MWR2904.1>.
- Gupta, V. K., and L. Duckstein, 1975: A stochastic analysis of extreme droughts. *Water Resour. Res.*, **11**, 221–228, <https://doi.org/10.1029/WR011i002p00221>.
- Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, **134**, 3209–3229, <https://doi.org/10.1175/MWR3237.1>.
- , —, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, [https://doi.org/10.1175/1520-0493\(2004\)132<1434:ERIMFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2).
- Herr, H. D., and R. Krzysztofowicz, 2005: Generic probability distribution of rainfall in space: The bivariate model. *J. Hydrol.*, **306**, 234–263, <https://doi.org/10.1016/j.jhydrol.2004.09.011>.
- Hersbach, R., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Kalbfleisch, J. D., and R. L. Prentice, 1980: *The Statistical Analysis of Failure Time*. Wiley and Sons, 435 pp.
- Katz, R. W., 1977: Precipitation as a chain-dependent process. *J. Appl. Meteor.*, **16**, 671–676, [https://doi.org/10.1175/1520-0450\(1977\)016<0671:PAACDP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1977)016<0671:PAACDP>2.0.CO;2).
- Kavvas, M. L., and J. W. Delleur, 1981: A stochastic cluster model of daily rainfall sequences. *Water Resour. Res.*, **17**, 1151–1160, <https://doi.org/10.1029/WR017i004p01151>.
- Kelly, K. S., and R. Krzysztofowicz, 2000: Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.*, **36**, 2643–2653, <https://doi.org/10.1029/2000WR900061>.
- Kotecha, J. H., and P. E. Djurić, 1999: Gibbs sampling approach for generation of truncated multivariate gaussian random variables. *Proc. 1999 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP99)*, Phoenix, AZ, IEEE, 2643–2653, <https://doi.org/10.1109/ICASSP.1999.756335>.
- Krzysztofowicz, R., 1992: Bayesian correlation score: A utilitarian measure of forecast skill. *Mon. Wea. Rev.*, **120**, 208–219, [https://doi.org/10.1175/1520-0493\(1992\)120<0208:BCSAUM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0208:BCSAUM>2.0.CO;2).
- , 1999: Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.*, **35**, 2739–2750, <https://doi.org/10.1029/1999WR900099>.
- Li, Y., and S. K. Ghosh, 2015: Efficient sampling methods for truncated multivariate normal and student-t distributions subject to linear inequality constraints. *J. Stat. Theory Pract.*, **9**, 712–732, <https://doi.org/10.1080/15598608.2014.996690>.
- Little, R. J. A., and D. B. Rubin, 2002: *Statistical Analysis with Missing Data*. Wiley Interscience, 409 pp.
- Mardia, K. V., 1970: Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530, <https://doi.org/10.1093/biomet/57.3.519>.
- , J. T. Kent, and J. M. Bibby, 1979: *Multivariate Analysis. Probability and Mathematical Statistics*. Academic Press, 512 pp.
- Matheson, J. E., and R. L. Winkler, 1976: Scoring rules for continuous probability distributions. *Manage. Sci.*, **22**, 1087–1096, <https://doi.org/10.1287/mnsc.22.10.1087>.
- Moran, P. A. P., 1970: Simulation and evaluation of complex water systems operation. *Water Resour. Res.*, **6**, 1737–1742, <https://doi.org/10.1029/WR006i006p01737>.
- Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338, [https://doi.org/10.1175/1520-0493\(1987\)115<1330:AGFFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFV>2.0.CO;2).
- Raftery, A. E., and S. Lewis, 1992: How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*, J. M. Bernardo et al., Eds., Oxford University Press, 763–773.
- , T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian Model Averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, <https://doi.org/10.1175/MWR2906.1>.
- Reggiani, P., G. Coccia, and B. Mukhopadhyay, 2016: Predictive uncertainty estimation on a precipitation and temperature reanalysis ensemble for Shigar basin, Central Karakoram. *Water*, **8** (6), 263, <https://doi.org/10.3390/w8060263>.
- , A. Boyko, T. Rientjes, and A. Khan, 2019: Probabilistic precipitation analysis in the Central Indus River basin. *Indus River Basin: Water Security and Sustainability*, S. Khan and T. Adams, Eds., Elsevier, 485 pp.
- Scheuerer, M., and T. M. Hamill, 2015: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon. Wea. Rev.*, **143**, 4578–4596, <https://doi.org/10.1175/MWR-D-15-0061.1>.
- Seo, D.-J., S. Perica, E. Welles, and J. Schaake, 2000: Simulation of precipitation fields from probabilistic quantitative precipitation forecast. *J. Hydrol.*, **239**, 203–229, [https://doi.org/10.1016/S0022-1694\(00\)00345-0](https://doi.org/10.1016/S0022-1694(00)00345-0).
- Sklar, A., 1959: Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, **1**, 229–231.
- Sloughter, J. M. L., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209–3220, <https://doi.org/10.1175/MWR3441.1>.
- Sorensen, D. A., D. Gianola, and I. R. Korsgaard, 1998: Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications. *Acta Agric. Scand. Sec. A, Animal Sci.*, **48**, 222–229, <https://doi.org/10.1080/09064709809362424>.
- Tanner, M. A., and W. H. Wong, 1987: The calculation of posterior distributions by data augmentation. *J. Amer. Stat. Assoc.*, **82**, 528–540, <https://doi.org/10.1080/01621459.1987.10478458>.
- Todini, E., 2008: A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manage.*, **36**, 3265–3277, <https://doi.org/10.1080/15715124.2008.9635342>.
- , and M. Di Bacco, 1997: A combined Pólya process and mixture distribution approach to rainfall modelling.

- Hydrol. Earth Syst. Sci.*, **1**, 367–378, <https://doi.org/10.5194/hess-1-367-1997>.
- , and F. Pellegrini, 1999: A maximum likelihood estimator for semi-variogram parameters in kriging. *GeoENVII—Geostatistics for Environmental Applications*, J. Gomez-Hernandez, A. Soares, and R. Froidevaux, Eds., Kluwer Academic Publishers, 187–198.
- Todorovic, P., and V. Yevjevich, 1969: Stochastic processes of precipitation. Colorado State University Hydrology Paper 35, 61 pp.
- Vrac, M., and P. Naveau, 2007: Stochastic downscaling of precipitations: From dry events to heavy rainfalls. *Water Resour. Res.*, **43**, W07402, <https://doi.org/10.1029/2006WR005308>.
- Wang, C. S., D. E. Robertson, and D. Gianola, 2011: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resour. Res.*, **47**, W02546, <https://doi.org/10.1029/2010WR009333>.
- Waymire, E., and V. K. Gupta, 1981: The mathematical structure of rainfall representations: 1. A review of the stochastic rainfall models. *Water Resour. Res.*, **17**, 1261–1272, <https://doi.org/10.1029/WR017i005p01261>.
- Wilks, D. S., 1990: Maximum likelihood estimation for the gamma distribution using data containing zeros. *J. Climate*, **3**, 1495–1501, [https://doi.org/10.1175/1520-0442\(1990\)003<1495:MLEFTG>2.0.CO;2](https://doi.org/10.1175/1520-0442(1990)003<1495:MLEFTG>2.0.CO;2).
- , 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. International Geophysics Series, Vol. 59, Elsevier, 467 pp.
- Woolhiser, D. A., and G. G. S. Pegram, 1979: Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. *J. Appl. Meteor.*, **18**, 34–42, [https://doi.org/10.1175/1520-0450\(1979\)018<0034:MLEOFC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<0034:MLEOFC>2.0.CO;2).