



## OPEN ACCESS

## EDITED BY

Georgia A. Papacharalampous,  
National Technical University of Athens,  
Greece

## REVIEWED BY

Patrick Enda O'Connell,  
Newcastle University, United Kingdom  
Jasper Vrugt,  
University of California, Irvine, United States

## \*CORRESPONDENCE

Paolo Reggiani  
✉ paolo.reggiani@uni-siegen.de

RECEIVED 21 December 2023

ACCEPTED 11 March 2024

PUBLISHED 16 April 2024

## CITATION

Reggiani P, Biondi D and Todini E (2024) On time-horizons based post-processing of flow forecasts. *Front. Water* 6:1359750. doi: 10.3389/frwa.2024.1359750

## COPYRIGHT

© 2024 Reggiani, Biondi and Todini. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# On time-horizons based post-processing of flow forecasts

Paolo Reggiani<sup>1,2\*</sup>, Daniela Biondi<sup>2,3</sup> and Ezio Todini<sup>2</sup>

<sup>1</sup>Department of Civil Engineering, Research Institute for Water and Environment, University of Siegen, Siegen, Germany, <sup>2</sup>Italian Hydrological Society, Bologna, Italy, <sup>3</sup>Department of Informatics, Modeling, Electronics and Systems Engineering, University of Calabria, Rende, Italy

Post-processing raw stream flow forecasts are generally understood as estimating the univariate predictive density of stage or discharge values at selected future time steps, which is conditional on a single or multiple streamflow forecasts and observations up to the forecast start time  $t_0$ . The predictive density indicates to a forecaster in the most comprehensive way which flood level is likely to be expected. To this end, a variety of post-processing methods were proposed, which have respective strengths and weaknesses. These methods focus near-exclusively on the probabilistic forecast of the predictand at a single set future time  $t_i$ , without addressing the predictive capability over the sequence of temporal sub-horizons  $(t_0, t_1] \subset (t_0, t_2] \subset \dots \subset (t_0, t_k]$  nested into the overall forecast horizon. Here, we demonstrate the advantages of time-horizon dependent processing of streamflow forecasts, which evaluates the evolution of the predictive density over the sub-horizons by considering the temporal correlation among forecast ensemble members in addition to their cross-correlation with observations. The resulting probabilistic forecast consists of a multivariate distribution of stages and/or discharges at lagged forecasting times. These multivariate predictive distributions have the advantage of providing the likelihood of exceeding a critical threshold during the forecasting horizon while simultaneously offering valuable insights into the expected time of such exceedance. This approach supports not only decisions on issuing timely flood warnings but also the planing and roll-out of mitigating actions.

## KEYWORDS

streamflow prediction, ensemble forecasting, post-processing, predictive uncertainty, multi-horizon, time correlation

## 1 Introduction

Operational streamflow forecasting is an effective non-structural measure to contain flood risk and protect human lives. Starting from weather models, which prognosticate future precipitation to drive stream flow models, one can generate a forecast of future river flow rates and/or stages. Forecasts of this type are non-deterministic, notwithstanding Earth-system models, and hydrological simulation tools rest on the solution of deterministic differential equations. A wide range of sources for uncertainty affect the forecast and need educated quantification, before being communicated to decision-makers (Krzysztofowicz, 2001a; Pagano et al., 2014). Streamflow uncertainty originates mainly from forecasted meteorological forcing and the chaotic nature of the atmospheric system, but equally from model inadequacies because of the coarse process parameterization and space-time discretization. Measurement errors play a role as well. The sources of uncertainty and its diffusion through the modeling system have been amply investigated in the literature, and comprehensive approaches have been implemented for forecast error correction and control, among which sequential data assimilation

of observations into models (Kitanidis and Bras, 1980; Seo et al., 2003; Vrugt et al., 2005; Kalnay et al., 2007; Clark et al., 2008). Nevertheless, the residual and irreducible uncertainty, which remains in each forecast, needs to be assessed in probabilistic terms through the predictive density, which expresses the uncertainty of the yet to be observed predictand, conditional on the predictors (Hamill and Whitaker, 2006).

The theoretical basis for a Bayesian forecasting system (BFS) has been described in a seminal article by Krzysztofowicz (1999), which lays out the structure of a probabilistic processor of uncertainties. The precipitation and flow/river stage random processes are elaborated by an independent precipitation uncertainty processor (PUP; Kelly and Krzysztofowicz, 2000) and a hydrological uncertainty processor (HUP; Krzysztofowicz and Kelly, 2000), respectively, that are combined through an integrator of densities into a probabilistic river stage forecast (Krzysztofowicz, 2001b). The dichotomy into PUP and HUP allows for separate processing of the meteorological uncertainty and the one inherent to the deterministic rainfall-runoff modeling process. The resulting river-stage forecasts consist of a sequence of univariate predictive probability distribution functions  $f(h_t|h_o, u_o, w_o)$  of water levels  $h_t$  for each forecasting time step  $t = 1, \dots, k$ , following forecast onset time  $t_o$ , conditional on the initial water level  $h_o$ , on hydrologic states  $u_o$ , and on a probabilistic quantitative precipitation forecast  $w_o$  at  $t_o$ . The cumulative distribution function associated with the density function  $f$  is  $F(h^*) = P(h_t < h^*|h_o, u_o, w_o)$ . The proposed processor structure is Bayesian as it rests on the convolution of a prior probability density of future streamflow with a likelihood function. The likelihood statistically specifies the error structure of model forecasts. One of the fundamental assumptions underlying the HUP is modeling the prior distribution of flow forecasts as a lag-one Markov process. The BFS has been shown to fulfill three general requirements for the reliability of a forecasting system (Krzysztofowicz, 1999):

- It quantifies all sources of uncertainty pertaining to the predictand, i.e., the stage or discharge forecast;
- It possesses a self-calibration property, wherein, in the long run, the probabilistic forecasts preserve the prior (historic) distribution of the predictand;
- It possesses a coherence property, wherein the economic value of the forecast is at least zero or positive, relative to the economic value of the prior distribution of the predictand.

The Bayesian HUP was applied to the River Rhine forecasting system for a single deterministic forecast (Reggiani and Weerts, 2008), whereas an ensemble application for the same system was presented by Reggiani et al. (2009).

In further steps, the BFS was elaborated into a time-horizon dependent processor with multivariate output. In contrast to the earlier version, the system now issues a multivariate probabilistic river stage forecast obtained on the basis of a model of flow water level changes represented by Markov lag-one transition densities. The probabilistic water stage forecast consists of a joint predictive probability density function  $f(h_1, \dots, h_k|h_o, u_o, w_o)$  with corresponding joint predictive distribution function  $P(h^*) = P(h_1 \leq h^*, \dots, h_k \leq h^*)$  defined through the multiple integral of  $f(\cdot|\cdot)$ .

By setting  $h_{max,k} = \max\{h_1, \dots, h_k\}$ , the maximum stage reached over the  $k$ -step forecast horizon of length  $t_k - t_o$ , function  $P(h^*)$  expresses the probability of at least one exceedance of the threshold value  $h^*$  across the  $k$  consecutive steps:

$$\begin{aligned} P(h_{max,k} > h^*) &= 1 - P(h_{max,k} \leq h^*) \\ &= 1 - P(h_1 \leq h^*, \dots, h_k \leq h^*) = 1 - P(h^*) \end{aligned} \quad (1)$$

meaning that  $P(h^*)$  is equivalent to the joint non-exceedance probability of the threshold value over the range of temporal horizons. One also needs to note that the joint distribution  $P(h^*)$  is indirectly defined over a sequence of nested temporal sub-horizons  $(t_o, t_1] \subset (t_o, t_2] \subset \dots \subset (t_o, t_k]$ , and therefore also allows to determine the probability of the time needed for threshold exceedance to occur. The probability on the time to reach flooding through exceedance of the threshold  $h^*$ ,  $t = t^*(h^*)$ , therefore, is:

$$P(t_k \geq t^*) = 1 - P(t_k < t^*) = 1 - P(h^*) \quad (2)$$

Capturing the uncertainty of the time to threshold exceedance is relevant in an operational setting, as it allows for adaptive decision-making (Krzysztofowicz, 2014). The original HUP was therefore adapted into a multivariate precipitation-dependent processor, which estimates the joint distribution of river stages, given stage and hydrological initial conditions, and a probabilistic precipitation forecast (Maranzano and Krzysztofowicz, 2004).

Ensemble prediction systems (EPS), in which multiple forecasts are generated under the assumption that their inherent spread describes the compound uncertainty of natural phenomena at different system stages, as well as the uncertainty due to meteorological and hydrologic model deficiencies, have become common in the last two decades in virtue of enhanced computing capabilities. An ensemble forecast in principle requires that a true random sample of the predictand is sufficiently large to allow for an adequate approximation of its potential probability distribution (Herr and Krzysztofowicz, 2015).

Forecast ensembles can be generated, for instance, from historical data using a single modeling system, which is initialized with perturbed initial condition vectors, or by combination of multiple models. Data-based ensembles on the other hand, which can be generated from reshuffling historical series (Wood and Schaake, 2008), suffer from the shortcoming that resulting ensembles are factually small due to strong cross-correlations among its members, while the non-stationarity of the process cannot be acknowledged (Scheffik et al., 2013). For model-generated ensembles, different processing methods have been devised for statistical post-adjustment. Some authors propose dressing single ensemble members (Pagano et al., 2013; Verkade et al., 2013, 2017), while methods such as the Bayesian HUP (Reggiani et al., 2009), the non-parametric processor by Brown and Seo (2010), process the whole ensemble of different model outputs all at once. The listed ensemble forecasting systems have different strengths and weaknesses, most do not output ensembles of time series of the predictand (Herr and Krzysztofowicz, 2015) but the predictive distribution of the predictand instead. In addition, the ensemble may not pertain to one and the same distribution, from which nature would sample the predictand and only few

of the EPS satisfy the requirements of calibration and coherence mentioned above.

Time-horizon dependent post-processing of streamflow ensembles using the ensemble model output statistics (EMOS) technique (Gneiting et al., 2005) has been introduced by Hemri et al. (2013) using multivariate Bayesian model averaging (Raftery et al., 2005). Objective of that study was to extend univariate predictive densities for individual lead times to jointly process the EMOS output for the entire forecasting horizon encompassing multiple lead time sub-horizons. The authors further elaborated time-horizon dependent post-processing by introducing Ensemble Copula Coupling (ECC) and Gaussian Copula Approach (GCA) for joining univariate single lead-time marginal EMOS output distributions with a multivariate time-predictive density covering the whole range of lead-times (Hemri et al., 2015; Hemri and Klein, 2017). The output can be expressed as quantile trajectories, which account for the correlations structure of the ensemble prediction over the forecast time horizon using a correlogram. The proposed approach has shown to perform well using classical performance indicators such as CRPS (Hersbach, 2000) in terms of calibration and sharpness. In addition to guaranteeing calibration, the proposed approach remains not necessarily self-calibrating or coherent.

The model-conditional processor (MCP) of uncertainty (Todini, 2008) is based on the conditional distribution of future occurrences given single or ensemble forecasts. As clearly emerged from Equations (14, 15), the weights multiplying the members of the forecast ensemble depend directly on the covariance of each member with observations and inversely on the increasing mutual covariance. This expresses the fact that additional information must be independent from the one already available for producing a gain in knowledge. After standardizing predictor and predictand, one estimates their mutual variance-covariance matrix. The elements of the matrix, actually Pearson correlations in the standard Gaussian space, represent weights that attribute different importance to members of the ensemble of predictors based on their statistical similarity with the predictand over a control period. Because this approach requires the joint analysis of the covariance structure between predictors and predictand from data samples, it becomes feasible to process a whole ensemble of predictors all at once. The MCP has been developed on the assumption that predictor and predictands can be transformed into standard Gaussian variates, and their mutual dependency structure approximated as a Gaussian copula. Once the dependency structure has been established, conditional densities can be derived analytically and returned into the space of origin. An application of the MCP to streamflow ensembles was presented by Coccia and Todini (2011). This application also emphasizes the challenge posed by the heteroscedastic of data, which has been solved by piecewise modeling the region of the multivariate distribution by truncated Gaussian copulas for sub-regions with near-homoscedastic dependency. The equivalence between MCP and the univariate Bayesian HUP (Krzysztofowicz and Kelly, 2000) was demonstrated for the case, in which one employs either a Markov lag-one prior distribution or a simple linear regression model between random samples of observations and past predictions (Todini, 2012). This inter-comparison also

showed that the Bayesian forecasting system concept (BFS), albeit theoretically consistent and analytically tractable, runs into practical limitations once multiple predictors become involved. Precisely, this drawback of the BFS has been the main driver for employing the MCP to estimate the predictive density of precipitation (Reggiani and Boyko, 2019) in operational weather forecasting, where up to 50 member ensembles (Toth and Kalnay, 1993; Molteni et al., 1996) are routinely used to approximate the distribution of the predictand. This study also provides an example on how imputation of missing data can be used to artificially extend left-bound samples of a censored variate, therefore enabling the application of the processor to an intermittent random process such as precipitation. Barbeta et al. (2017) introduced the joint probability density concept over a nested set of time horizons into the MCP and conditioned the predictive density on an ensemble of streamflow forecasts. The aim of this study is to revisit the time-horizon dependent MCP with the following specific research objectives:

- Explain in more detail the stochastic dependency structure and corresponding derivations for the time-horizon dependent case,
- Demonstrate adherence to conditions of self-calibration and coherence,
- Discuss the advantages of time-horizon dependency in forecast processing over time-horizon independence,
- Show how the output of the uncertainty processor can be elaborated into forecasting products for advanced decision-making in operational settings (Pagano et al., 2014).

To pursue these objectives, the article is structured in four sections. The first section describes the methods, the second presents a numerical application, whereas the third section is devoted to discussing the approach. The last section concludes.

## 2 Methods

### 2.1 Predictive density

The predictive density is defined as the conditional probability density function (PDF) of an unknown quantity, in our case, stage or discharge; given all presently available information on that quantity, here the ensemble of streamflow estimates produced by a deterministic hydraulic model (Hamill and Whitaker, 2006) is as follows:

$$f(\mathbf{h}|\hat{\mathbf{h}}) \quad (3)$$

where  $\mathbf{h}$  is the vector of values of the predictand at times  $t = 1, \dots, n$  and  $\hat{\mathbf{h}}$  is an array containing an  $m$ -member ensemble of synchronous predictors of  $\mathbf{h}$ , which will be specified in more detail in Section 2.3. These become the known quantities, which enhance our knowledge by conditioning the probability distribution of the observations  $\mathbf{h}$ . The predictors  $\hat{\mathbf{h}}$  act as error-affected estimates of  $\mathbf{h}$  and serve as knowledge support in reducing uncertainty through bias-removal and sharpening of the predictive density (Equation 3) (Wilks, 1995). The predictive

distribution can be obtained in different ways, for instance, through Quantile Regression (Koenker, 2005), the Bayesian hydrologic uncertainty processor (Krzysztofowicz and Kelly, 2000), Bayesian model averaging (Raftery et al., 2005), or the model-conditional processor (Todini, 2008). All these approaches have different strengths and weaknesses. The main reasons for adopting the MCP here is the ease with which it allows to include an ensemble of  $m$  predictors, while it is much more cumbersome using alternate methods. Unlike BMA, it is not restricted by the assumption of independent forecast errors, which one cannot make when studying an auto-correlated process.

## 2.2 Probabilistic flow forecast

A streamflow ensemble forecast  $\hat{\mathbf{h}}$  of the not yet observed predictand  $\mathbf{h}$  is generated using a deterministic hydrological model  $\hat{\mathbf{h}} = M_h(\mathbf{h}_o, \mathbf{w}, \mathbf{u})$ , which is driven by a meteorological forcing vector  $\mathbf{w}$  and depends on past observations of  $\mathbf{h}_o$  and a vector of hydrological states  $\mathbf{u}$ . The weather forecast  $\hat{\mathbf{w}}$  output by a weather model  $\hat{\mathbf{w}} = M_w(\mathbf{w}_o, \mathbf{v})$  induces a density  $\pi(\hat{\mathbf{w}}|\mathbf{w}_o, \mathbf{v})$  conditional on a vector of atmospheric states  $\mathbf{v}$  and past  $\mathbf{w}_o$ .  $M_h(\mathbf{h}_o, \mathbf{w}, \mathbf{u})$  together with  $\pi(\hat{\mathbf{w}}|\mathbf{w}_o, \mathbf{v})$  induce a density  $\gamma(\hat{\mathbf{h}}|\mathbf{h}_o, \mathbf{w}, \mathbf{u})$  of  $\hat{\mathbf{h}}$ . Densities  $f(\mathbf{h}|\hat{\mathbf{h}})$  and  $\xi(\mathbf{w}|\hat{\mathbf{w}})$  assess the predictive uncertainty (Equation 3) by processing outputs of models  $M_w$  and  $M_h$ . The role of the processor is to de-bias and variance-adjust model forecasts so that the corrected output distributions better match those of the unconditional long-term distributions of observed streamflows. The probabilistic flow forecast is produced by integrating hydrologic and input uncertainty according to the law of total probability:

$$\psi(\mathbf{h}|\mathbf{h}_o, \mathbf{w}_o, \mathbf{u}, \mathbf{v}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{h}|\hat{\mathbf{h}}) \gamma(\hat{\mathbf{h}}|\mathbf{h}_o, \mathbf{w}, \mathbf{u}) \xi(\mathbf{w}|\hat{\mathbf{w}}) \pi(\hat{\mathbf{w}}|\mathbf{w}_o, \mathbf{v}) d\hat{\mathbf{h}} d\mathbf{w} d\hat{\mathbf{w}} \quad (4)$$

in which the dependency on model outputs of weather and stage/discharge has been marginalized out. In doing so, we assume that an output ensemble constitutes a sufficiently sized sample (Herr and Krzysztofowicz, 2010) for approximating the true densities of  $\hat{\mathbf{w}}$  and  $\hat{\mathbf{h}}$ . The probabilistic forecast density  $\psi(\mathbf{h}|\mathbf{h}_o, \mathbf{w}_o, \mathbf{u}, \mathbf{v})$  must be coherent in the sense that it is derivable from a super-density  $\zeta(\mathbf{h}, \mathbf{h}_o, \mathbf{w}_o, \mathbf{u}, \mathbf{v})$  which characterizes the joint natural weather and streamflow phenomena. This requires the probabilistic flow forecast to be well-calibrated, i.e., truly representative of the overall natural phenomenon. This means that marginal densities of selected predicted process variables must fall back onto those of historical observations. In particular, for probabilistic flow or stage forecasts, this equates to Krzysztofowicz (1999):

$$E_{\mathbf{w}, \mathbf{u}, \mathbf{v}}[\psi(\mathbf{h}|\mathbf{h}_o, \mathbf{w}_o, \mathbf{u}, \mathbf{v})] = g(\mathbf{h}) \quad (5)$$

where  $g(\mathbf{h})$  is the unconditional probability density of observations. Let us for the purpose of this study suppose the weather to be fully known or a representative ensemble drawn from a perfectly calibrated predictive density  $\xi(\mathbf{w}|\hat{\mathbf{w}})$ . This sample is then used to generate a streamflow prediction ensemble  $\hat{\mathbf{h}}$ , which is a realization

from the distribution  $\gamma(\hat{\mathbf{h}}|\mathbf{h}_o, \mathbf{w}, \mathbf{u})$  for set  $\mathbf{w}$  and  $\mathbf{u}$  and fully specifies the hydrological uncertainty attributable to the model imperfections and sub-optimal knowledge of system states. From here onward, we focus exclusively on the predictive density  $f(\mathbf{h}|\hat{\mathbf{h}})$  for streamflow with a time-horizon dependency, while weather uncertainty is assumed to be perfectly specified and transferred into the density  $\gamma$ , from which we draw the ensemble  $\hat{\mathbf{h}}$  of streamflow predictions. Equation (4) of the probabilistic river flow forecast clearly separates relevant sources of uncertainty.

## 2.3 Time-horizon dependency

The multivariate MCP is a processor of uncertainty, which is based on mapping variates into standard Gaussian ones and assuming a linear dependency structure modeled in the Gaussian space. For this reason, it can also be called meta-Gaussian processor. If  $\mathbf{h}$  is a continuous random variate describing a stochastic process such as stage or discharge, its predictor is denoted as  $\hat{\mathbf{h}}$ . Discrete realizations of the variate are represented by the following lower-case vectors:

$$\mathbf{h} = (h_1, \dots, h_n) \quad (6a)$$

$$\hat{\mathbf{h}} = [(\hat{h}_{1,1}, \dots, \hat{h}_{n,1}), \dots, (\hat{h}_{1,m}, \dots, \hat{h}_{n,m})] \quad (6b)$$

where  $1, \dots, n$  are the number of time steps at which realizations of  $\mathbf{h}$  occur and  $m$  is the number of synchronous predictors of  $\mathbf{h}$ , for instance, (multi-)model ensembles. The mapping of the sample realizations into their own standard normal images is denoted with  $\boldsymbol{\eta} = \Phi^{-1}(F(\mathbf{h}))$  and  $\hat{\boldsymbol{\eta}} = \Phi^{-1}(G(\hat{\mathbf{h}}))$  and respective realizations with  $\boldsymbol{\eta}$  and  $\hat{\boldsymbol{\eta}}$ . The mapping occurs by matching probabilities of the empirical CDF with a standard-Gaussian distribution, either via a non-parametric method referred to as Normal Quantile Transform (NQT) or by probability matching after fitting the classical parametric probability distribution functions.  $F$  and  $G$  denote the marginal distributions (empirical or parametric) and  $\Phi^{-1}$  the inverse standard normal CDF. After organizing observations and synchronous predictions by  $m$  models into  $k$  time-horizon bins of the forecasts,  $j = 1, \dots, k$ , indicating the  $j \cdot \Delta t$  time slice of the forecast horizon counting from  $t_o$ , we introduce Gaussian images of the predictor and predictand vectors in Equation (6):

$$\boldsymbol{\eta} = (\eta_{t_o+\Delta t}, \dots, \eta_{t_o+k\Delta t}) = (\eta_1, \dots, \eta_k) \quad (7a)$$

$$\hat{\boldsymbol{\eta}} = [(\hat{\eta}_{t_o+\Delta t,1}, \dots, \hat{\eta}_{t_o+k\Delta t,1}), \dots, (\hat{\eta}_{t_o+\Delta t,m}, \dots, \hat{\eta}_{t_o+k\Delta t,m})] = [(\hat{\eta}_{1,1}, \dots, \hat{\eta}_{k,1}), \dots, (\hat{\eta}_{1,m}, \dots, \hat{\eta}_{k,m})] \quad (7b)$$

Next, we assume that the Gaussian predictand  $\boldsymbol{\eta}$  relates to predictors  $\hat{\boldsymbol{\eta}}$  through the following linear model:

$$\boldsymbol{\eta} = \boldsymbol{\mu}_\eta + \mathbf{B} \cdot (\hat{\boldsymbol{\eta}} - \boldsymbol{\mu}_{\hat{\boldsymbol{\eta}}}) + \boldsymbol{\omega} \quad (8)$$

where  $\mathbf{B}$  is a  $k \times m \cdot k$  matrix of multi-linear regression coefficients,  $\boldsymbol{\mu}_\eta$  is a  $k \times 1$  vector of means  $E(\boldsymbol{\eta})$ ,  $\boldsymbol{\mu}_{\hat{\boldsymbol{\eta}}}$  an  $m \cdot k \times 1$  vector of means  $E(\hat{\boldsymbol{\eta}})$ , while  $\boldsymbol{\omega}$  is a  $k \times 1$  vector of errors with probability distribution  $N(\mathbf{0}, \Sigma_\omega)$ , typically a Gaussian noise, statistically independent of  $\boldsymbol{\eta}$  and  $\hat{\boldsymbol{\eta}}$ . The conditional mean is modeled as a multi-linear regression:

$$E(\boldsymbol{\eta}|\hat{\boldsymbol{\eta}}) = \boldsymbol{\mu}_{\boldsymbol{\eta}|\hat{\boldsymbol{\eta}}} = \boldsymbol{\mu}_\eta + \mathbf{B} \cdot (\hat{\boldsymbol{\eta}} - \boldsymbol{\mu}_{\hat{\boldsymbol{\eta}}}) \quad (9)$$

whereas the variance is conditional on  $\hat{\eta}$  as follows:

$$\text{Var}(\eta|\hat{\eta}) = \Sigma_{\eta|\hat{\eta}} = E[(\eta - \mu_{\eta|\hat{\eta}})(\eta - \mu_{\eta|\hat{\eta}})^T] \quad (10)$$

**Assumption:** The joint  $k \times (m + 1)$ -sized sample of NQT-transformed observations and predictions  $(\eta, \hat{\eta})$  is multivariate normal distributed (MVND):

$$(\eta, \hat{\eta}) \sim N_{k \cdot (m+1)}(\mu, \Sigma) \quad (11)$$

with density

$$\phi(\eta, \hat{\eta}) = \frac{\exp\left\{-\frac{1}{2}[(\eta, \hat{\eta}) - \mu]^T \Sigma^{-1} [(\eta, \hat{\eta}) - \mu]\right\}}{\sqrt{(2\pi)^{k \cdot (m+1)} \|\Sigma\|}} \quad (12)$$

where  $\Sigma$  is the covariance matrix of the joint sample  $(\eta, \hat{\eta})$  and  $\mu = E(\eta, \hat{\eta})$  the  $k \cdot (m + 1)$ -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 assigning a Gaussian copula dependence to the standard normal variates  $\eta$  and  $\hat{\eta}$ . If so, conditional densities can be obtained as follows:

$$\phi(\eta|\hat{\eta}) = \frac{\phi(\eta, \hat{\eta})}{\phi(\hat{\eta})} \sim N_{k \cdot (m+1)}(\mu_{\eta|\hat{\eta}}, \Sigma_{\eta|\hat{\eta}}) \quad (13)$$

with  $\phi(\hat{\eta})$  the  $1, \dots, m$  standard normal marginal densities of Equation (12). Equation (13) represents a family of standard normal predictive densities with conditional mean and variance:

$$\mu_{\eta|\hat{\eta}} = \mu_{\eta} + \Sigma_{\eta\hat{\eta}} \cdot \Sigma_{\hat{\eta}\hat{\eta}}^{-1} \cdot (\hat{\eta} - \mu_{\hat{\eta}}) \quad (14a)$$

$$\Sigma_{\eta|\hat{\eta}} = \Sigma_{\eta\eta} - \Sigma_{\eta\hat{\eta}} \cdot \Sigma_{\hat{\eta}\hat{\eta}}^{-1} \cdot \Sigma_{\hat{\eta}\eta} \quad (14b)$$

with  $\Sigma_{\eta\eta}$  the  $k$ -sized matrix of observations covariances,  $\Sigma_{\hat{\eta}\hat{\eta}}$  the  $k \cdot m$ -sized covariance matrix of  $\hat{\eta}$ , and  $\Sigma_{\eta\hat{\eta}}$  the  $k \cdot m \times k$  covariance matrix of  $\eta$  and  $\hat{\eta}$ .

Because observations  $\mathbf{h}$  as well as model forecasts  $\hat{\mathbf{h}}$  are mapped into standard Gaussian variables, i.e.,  $\eta$  and  $\hat{\eta} \sim N(0, 1)$ , the vectors of means  $\mu_{\eta} = \mu_{\hat{\eta}} = \mathbf{0}$ , whereas  $\Sigma_{\eta\eta}$  and  $\Sigma_{\hat{\eta}\hat{\eta}}$  have diagonal terms equal to unity variances. Equations (14) therefore simplify:

$$\mu_{\eta|\hat{\eta}} = \Sigma_{\eta\hat{\eta}} \cdot \Sigma_{\hat{\eta}\hat{\eta}}^{-1} \cdot \hat{\eta} \quad (15a)$$

$$\Sigma_{\eta|\hat{\eta}} = \Sigma_{\eta\eta} - \Sigma_{\eta\hat{\eta}} \cdot \Sigma_{\hat{\eta}\hat{\eta}}^{-1} \cdot \Sigma_{\hat{\eta}\eta} \quad (15b)$$

The adequacy of the meta-Gaussian model assumption is relatively easy to verify in the uni- or bivariate case, but less so with multiple variates and can only be assessed a posteriori by verifying the predictive distribution against the one of observations. The conventional case of a forecast for a single time horizon, for instance, when assessing the predictive density of one and only one future value  $\eta_{t_o+j\Delta t}$  given the predictions issued at time  $t_o$  by a single or  $m$  predictors, Equation (15) reduces to scalar conditional mean and variance of the Gaussian univariate predictive density for said forecast time:

$$\mu_{\eta|\hat{\eta}} = \sigma_{\eta\hat{\eta}} \cdot \Sigma_{\hat{\eta}\hat{\eta}}^{-1} \cdot \hat{\eta} \quad (16a)$$

$$\sigma_{\eta|\hat{\eta}}^2 = 1 - \sigma_{\eta\hat{\eta}} \cdot \Sigma_{\hat{\eta}\hat{\eta}}^{-1} \cdot \sigma_{\hat{\eta}\eta} \quad (16b)$$

where  $\sigma_{\eta\hat{\eta}}$  is the  $m \times 1$  covariance vector between the predictand  $\eta$  and the  $m$  predictors  $\hat{\eta}$ , whereas  $\Sigma_{\hat{\eta}\hat{\eta}}$  is the  $m \times m$  covariance matrix of predictors.

The cumulative distribution function (CDF) of density (Equation 13) is obtained by multiple integration over the semi-infinite domain  $(-\infty, \hat{\eta}^*)$  with  $\hat{\eta}^*$  a single Gaussian threshold value common to all forecasting time steps:

$$\Phi(\eta_1 \leq \eta^*, \dots, \eta_k \leq \eta^*|\hat{\eta}) = \int_{-\infty}^{\eta^*} \dots \int_{-\infty}^{\eta^*} \phi(\eta|\hat{\eta}) d\eta \quad (17)$$

where once again the subscript  $j = 1, \dots, k$  abbreviates  $t_o + j\Delta t$ . The corresponding conditional CDF image in the real space is:

$$\begin{aligned} P(h_1 \leq h^* \dots h_k \leq h^*|\hat{\mathbf{h}}) \\ = \int_{-\infty}^{h^*} \dots \int_{-\infty}^{h^*} f(h_1, \dots, h_k | (\hat{h}_{1,1}, \dots, \hat{h}_{k,1}), \dots, (\hat{h}_{1,m}, \dots, \hat{h}_{k,m})) \\ dh_1 \dots dh_k \end{aligned} \quad (18)$$

A possible wider dependency  $f(\mathbf{h}|\hat{\mathbf{h}}, \mathbf{u}, \mathbf{w})$  on initial conditions and/or hydrological states or weather input as indicated in Equation (4) is omitted because it is irrelevant to the topic of time-horizon dependency.

The full multivariate conditional probability density  $\phi(\eta|\hat{\eta})$  given by Equation (13) is essential for determining the probability of threshold exceedance within the forecasting horizon and the expected time of occurrence, but there is no need for its full backtransformation into the real space. Instead, the single predictions and their predictive densities can be back-transformed into the real space using the inverse NQT via their marginal conditional distributions  $\phi(\eta_j|\hat{\eta}); \forall j = 1, \dots, k$ , easily obtained from the multivariate Gaussian conditional distribution.

## 2.4 Calibration and coherence

It can be shown that the predictive density (Equation 13) is a well-calibrated uncertainty assessor (Alpert and Raiffa, 1982; Krzysztofowicz, 1999). For a processor to be well-calibrated, the expectation  $E_{\hat{\eta}}[\phi(\eta|\hat{\eta})]$  must match the unconditioned historical distribution of streamflows when evaluated over many consecutive forecasting instances. The definition of conditional distribution applied to Equation (13) leads to:

$$E_{\hat{\eta}}[\phi(\eta|\hat{\eta})] = \phi(\eta) \quad (19)$$

and in the space of origin:

$$E_{\hat{\mathbf{h}}}[f(\mathbf{h}|\hat{\mathbf{h}})] = g(\mathbf{h}) = f(F^{-1}(\phi(\eta))) \quad (20)$$

In other words, when integrated over the entire space of possible predictions  $\hat{\mathbf{h}}$ , the predictive probability density function collapses onto the unconditional density of historic observations  $g(\mathbf{h})$ . As a corollary of this property, it follows that if the predictive information of  $\hat{\eta}$  is poor, i.e.,  $\Sigma_{\eta\hat{\eta}} \sim \mathbf{0}$  due to the predictions becoming fully uncorrelated with observations, the distribution of the forecast (Equation 13) is guaranteed to approach the prior density, and hence satisfies a property of coherence, meaning that the system guards the forecaster from issuing a forecast that is poorer than relying on statistics of past observations.

## 2.5 Advantages of time-horizon dependent flow forecasting

The extension of a classical single or multi-model probabilistic forecast to include time-horizon dependency bears the potential to considerably improve forecast accuracy, because it allows to account for the temporal evolution of the forecasting error. However, this is not the sole benefit of time-horizon dependency. The family of multivariate time-horizon dependent predictive densities (Equation 13) allows answering two important questions that are of particular interest and relevance to decision makers. In particular, when applying the method to the process  $\mathbf{h} = (h_1, \dots, h_n)$  and a triggering threshold  $h^*$ , associated with crossing a warning level or overtopping a levee by a future flow or water level, the following is of relevance:

- Which is the likelihood (or probability) that a threshold value  $h^*$  will be overtopped (or that flooding will occur) within the next  $j$  hours, when  $t_j - t_0, j = 1, \dots, k$  hours represents the length of the forecasting horizon of interest?
- At what time will threshold exceedance most likely occur?

To answer the first question on the probability of  $h$  overtopping  $h^*$  during the whole forecast horizon consisting of  $k$  intervals, one must start estimating the joint probability  $f(h_1, \dots, h_k | \hat{\mathbf{h}})$  of future water level occurrence at times  $j = 1, \dots, k$ ; conditional on predictor  $\hat{\mathbf{h}}$ , the  $k \times m$  vector of  $m$  model forecasts at  $k$  forecasting times defined in Equation (6b).

Once density  $f(h_1, \dots, h_k | \hat{\mathbf{h}})$  is specified, the probability of having at least one overtopping event in a generic time interval within the time horizon of interest equals one minus the probability of non-exceedance  $P(h_j < h^* | \hat{\mathbf{h}}) j = 1, \dots, k$ , across all nested sub-horizons of the forecast lead time  $(t_0, t_1) \subset (t_0, t_2) \subset \dots \subset (t_0, t_k)$ :

$$P(h > h^* | \hat{\mathbf{h}}) = 1 - \int_{-\infty}^{h^*} \dots \int_{-\infty}^{h^*} f(h_1, \dots, h_k | \hat{\mathbf{h}}) dh_1 \dots dh_k \quad (21)$$

In reply to the second question, Equation (21) must be explicitly stated in terms of the nested increasing forecasting sub-horizons:

$$P(h_j < h^* | \hat{\mathbf{h}}) = \int_{-\infty}^{h^*} \dots \int_{-\infty}^{h^*} f(h_1, \dots, h_j | \hat{h}_{1,1}, \dots, \hat{h}_{k,m}) dh_1 \dots dh_j \quad (22)$$

with  $P(h_j < h^* | \hat{\mathbf{h}})$  monotonously increasing with  $j$ . Equation (2) defined the distribution function of the time to flooding, which equals the probability of threshold water level exceedance, i.e.,  $P(t > t^* | \hat{\mathbf{h}}) = P(h(t) > h^* | \hat{\mathbf{h}})$ . Hence, the time rate of change of the exceedance probability equals the PDF of time to flooding:

$$\frac{dP(h(t) > h^* | \hat{\mathbf{h}})}{dt} = \frac{P(t > t^* | \hat{\mathbf{h}})}{dt} = f(t) \quad (23)$$

In discrete terms,  $f(t)$  can be approximate by finite probability increment ratios  $IR(t_0 + j\Delta t)$  defined for steps  $j = 1, \dots, k$ :

$$IR(t_0 + j\Delta t) = \frac{P(t_{j+1} > t^* | \hat{\mathbf{h}}) - P(t_j > t^* | \hat{\mathbf{h}})}{\Delta t} \quad (24)$$

The increment ratios express the time rate of change of the exceedance probability of the threshold as time elapses. Equation (24) can be used by a forecaster to identify the probability at what time the water level exceeds a preset threshold value  $h^*$  and allows for a vivid representation of the threshold exceedance probability, as will be shown through an example in Section 3.3.

## 3 Application

### 3.1 Data

To calibrate and validate the proposed processor, we use operational flow forecasts for the River Po, Italy, at gauging section Pontelagoscuro (PLS), which is considered the closure cross-section of the basin as shown in Figure 1. The Po River encompasses a drainage basin area of  $\sim 74,000$  km<sup>2</sup> in correspondence of the delta and is Italy's largest river system. Approximately 96% of the basin is situated within national borders, nearly one-third of which constitutes the alluvial plain of the River Po Valley. The remaining 4% lie mostly in Switzerland and, for very a small section, in France. The annual mean flow recorded at PLS is roughly 1,500 m<sup>3</sup>/s, with maxima  $\sim 10,000$  m<sup>3</sup>/s reached during extreme events. ARPA-SIM, the Hydro-Meteorological Service of the Emilia-Romagna Regional Agency for Environmental Protection, which operates the real-time River Po flood forecasting system, supplied hydrological streamflow ensemble forecasts based on Consortium for Small-Scale Modeling (COSMO) Limited Area Ensemble Prediction System (LEPS) ensemble members, which are used to force the process-based hydrological distributed model TOPKAPI (Liu et al., 2005). Streamflow routing toward gauging station PLS is performed by the Saint-Venant equation solver SOBEK (Stelling and Verwey, 2006). COSMO-LEPS is a joint development effort aimed at the improvement of short-to-medium range forecast of extreme and localized weather events in Europe. The system consists of 16 integrations of the COSMO weather model at 7-km spatial resolution that were launched on the basis of initial and boundary conditions provided by 16 representative members of the super-EPS (Molteni et al., 1996) of the European Centre for Medium-range Weather Forecasts (ECMWF).

Due to the provision of final streamflow forecasts at station PLS, it is not possible to separately quantify the uncertainties contributed by the hydrologic model from those induced by the hydraulic flood wave propagation model. However, Equation (4) indicates a formal framework for specifying and integrating different individual sources of uncertainty in the forecasting chain.

Observations and forecasts are sampled at 3-hourly time steps ( $\Delta t = 3$  h), the latter consisting of an array of hourly streamflow forecasts over a 120-h horizon. The available data set covers the period from 9/11/2012 to 28/02/2015, including 842 days and 1,628 forecasting instances in total.

### 3.2 Processing

As a first step in data processing, we organize all streamflow forecast outputted by the 16 member EPS ( $m = 16$ ) and

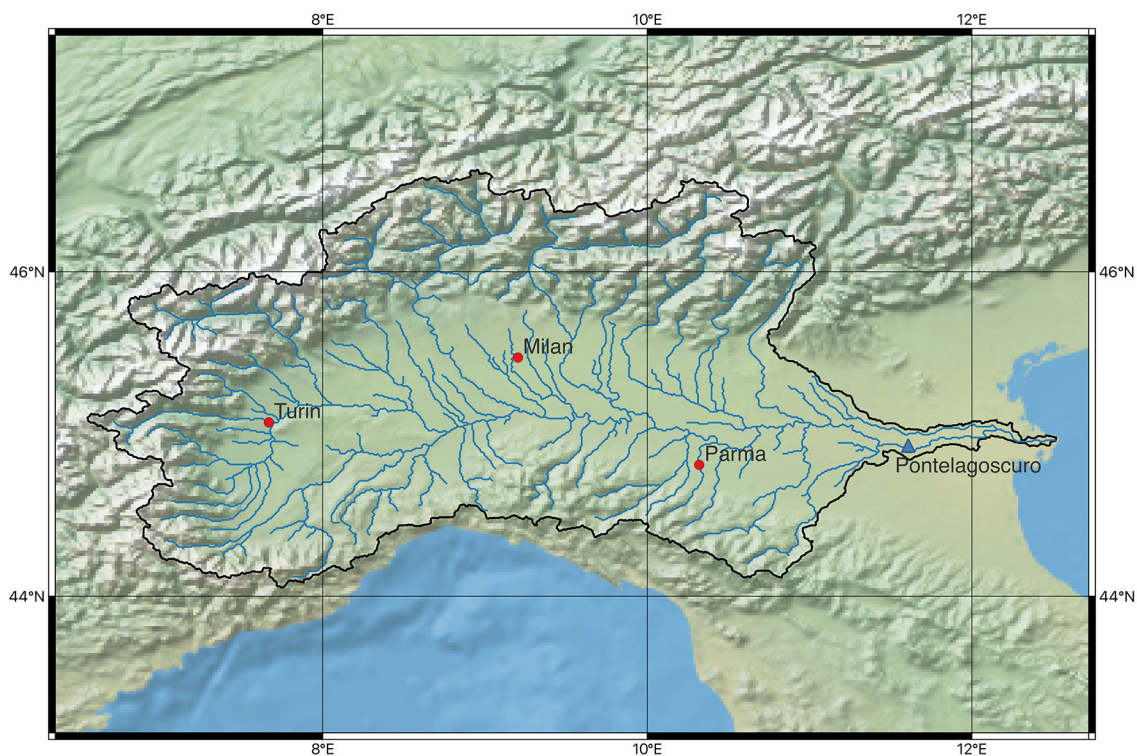


FIGURE 1  
River Po basin area. The location of station Pontelagoscuro (PLS) is indicated by a blue triangle.

observations for 1,628 forecasting instances such that each stream-flow forecast associated with an ensemble member compares directly against the corresponding observations. The sample of forecasts and associated observations over all selected  $k$  time steps in the 96 h forecasting horizon includes  $1,628 \times (m+1)$  values. Each sample pair  $(\mathbf{h}, \hat{\mathbf{h}})$  is mapped into its standard Gaussian image  $(\boldsymbol{\eta}, \hat{\boldsymbol{\eta}})$  by applying the Normal Quantile Transform (NQT). The mapping involves matching the probabilities of the empirical CDF of the original sample with those of the standard Gaussian distribution. In the standard Gaussian space, it is straightforward to establish the joint distribution (Equation 12) and the family of conditional densities (Equation 13). The latter yields  $k$  univariate predictive distributions, which are conditional on the  $k \cdot m$ -sized multivariate distribution of predictions. The image in the real space is obtained by mapping the  $k$  univariate marginal conditional distributions back into the space of origin by inverse NQT as  $\phi(\boldsymbol{\eta}_j | \hat{\boldsymbol{\eta}}) \rightarrow f(\mathbf{h}_j | \hat{\mathbf{h}}); j = 1, \dots, k$ .

### 3.3 Results

Next, we present the outcome of the processing in terms of performance indicators of the time-horizon dependent MPC (MCP-THD), whereby we compare the predictive mean against (a) the raw ensemble mean (REM) and (b) the predictive mean of the time-horizon independent MCP (MCP-THI). The time-horizon independent MCP-THI (also labeled multivariate MCP; Coccia and Todini, 2011) was extensively compared by Biondi

and Todini (2018) against the post-processing of forecasts using Bayesian model averaging (BMA; Raftery et al., 2005) and a univariate version of MCP. This univariate version is characterized by reduced mathematical complexity and can be convenient for applications where hydrological ensembles are provided; it considers the ensemble mean as the predictor but does not require the reordering of the ensemble members. This research demonstrates the advantage of using post-processors over direct use of raw ensembles in terms of accuracy and reliability. It also highlights, among other factors, the effects of ensemble member cross-correlation, which, unlike BMA and univariate MCP, is fully acknowledged by MCP-THI and MCP-THD.

We inter-compare REM, MCP-THI, and MCP-THD performances by executing a continuous forecast simulation over the 842-day simulation period indicated in Paragraph 3.1 and generating 1,628 forecast instances. At each of the  $k = 32$  3-hourly time steps of the 96 h forecast horizon, we compute the Nash-Sutcliffe efficiency (NSE) and root mean square error (RMSE) performance indicators. Both indicator values, which are listed in Table 1 for increasing time horizons, clearly indicate an improvement of performance when moving from REM to MCP-THI and finally to MCP-THD. This is also made visible in Figure 2, wherein the left pane shows the decline of NSE for the three processing methods with increasing lead time, whereas the progressive forecast error growth measured in terms of RMSE is given to the right.

Figure 3 shows the evolution of the discharge forecast for four different onset times  $t_o$ , which approach a selected flood event on

TABLE 1 Performance indicators for REM and the predictive means of MCP-THI and MCP-THD, 96 h lead time,  $\Delta t = 3$  h, shown every 6 h: Nash-Sutcliffe efficiency (NSE) and root mean square error (RMSE); the computation is carried out for the entire database including 1,628 forecasting instances.

Lead time (h)	REM		MCP-THI		MCP-THD	
	NSE (-)	RMSE ( $m^3/s$ )	NSE (-)	RMSE ( $m^3/s$ )	(-)	RMSE ( $m^3/s$ )
6	0.999	45.023	0.999	43.331	0.999	39.906
12	0.995	85.141	0.995	84.053	0.996	73.799
18	0.990	118.588	0.991	114.798	0.993	102.684
24	0.985	148.803	0.985	144.278	0.988	130.202
30	0.978	175.937	0.979	171.972	0.983	155.221
36	0.972	202.215	0.973	197.034	0.978	178.632
42	0.964	225.867	0.967	218.584	0.972	198.840
48	0.958	245.931	0.961	235.659	0.968	214.510
54	0.952	263.015	0.955	253.840	0.964	228.557
60	0.946	278.971	0.95	268.848	0.959	242.608
66	0.941	291.930	0.946	279.178	0.955	253.670
72	0.936	302.661	0.941	290.514	0.951	264.435
78	0.932	313.079	0.937	301.181	0.947	276.194
84	0.927	323.689	0.933	309.840	0.942	288.229
90	0.922	335.482	0.929	318.753	0.938	299.825
96	0.917	344.795	0.925	327.886	0.933	310.835

the 29/11/2012 in 3-hourly time steps. The shaded areas indicate the 90% confidence intervals tracked by the predictive distribution, and the solid black line are observations. The predictive mean, i.e., the expected value of the predictive distribution at different forecasting times, is shown by a dashed red line, whereas the  $4,000 m^3/s$  flood warning threshold is represented by the blue horizontal line.

The very essence of the MCP-THD processor is summarized in Figure 4 for the selected flood event of Figure 3 and chosen forecast onset time  $t_o = 29/11/2012$  09:00. The upper pane shows the forecast and respective 90% confidence intervals against observations. The middle pane indicates the cumulative probability density function (CDF) of threshold exceedance with changing forecasting horizon  $j = 1, \dots, k$  as defined by Equation (22). The lower pane shows the discrete increment ratios  $IR(t_o + j\Delta t)$  defined in Equation (24) for the same temporal increments, which equals the discrete PDF associated with the CDF. As one would expect, the probability mass is concentrated around time horizon  $t = 45$  h, when threshold exceedance actually occurs.

Next, we examine the forecast skill of the time-horizon dependent processor against the remaining two processing methods by comparing performance indicators. These include the Threat Score (TS), Probability of Detection (POD), and False Alarm Rate (FAR). The definition of the three indicators is recalled in Appendix A. For this purpose, an event is considered each time a flow value exceeds a predefined warning threshold (e.g.,  $4,000 m^3/s$ ) within the forecast horizon (from 1 to 96 h following the forecast onset time  $t_o$ ), and performance indices are consequently estimated. If we consider as flood event the one in which the peak flow exceeds the threshold value, there have been nine recorded

events with peak values exceeding  $4,000 m^3/s$  during the analyzed period. Just like any model, the forecast's robustness depends on the diversity, representativeness and completeness of the training and calibration data. A greater number of events can enable the model to better understand complex relationships in the data, enhancing its ability to make more accurate and reliable predictions.

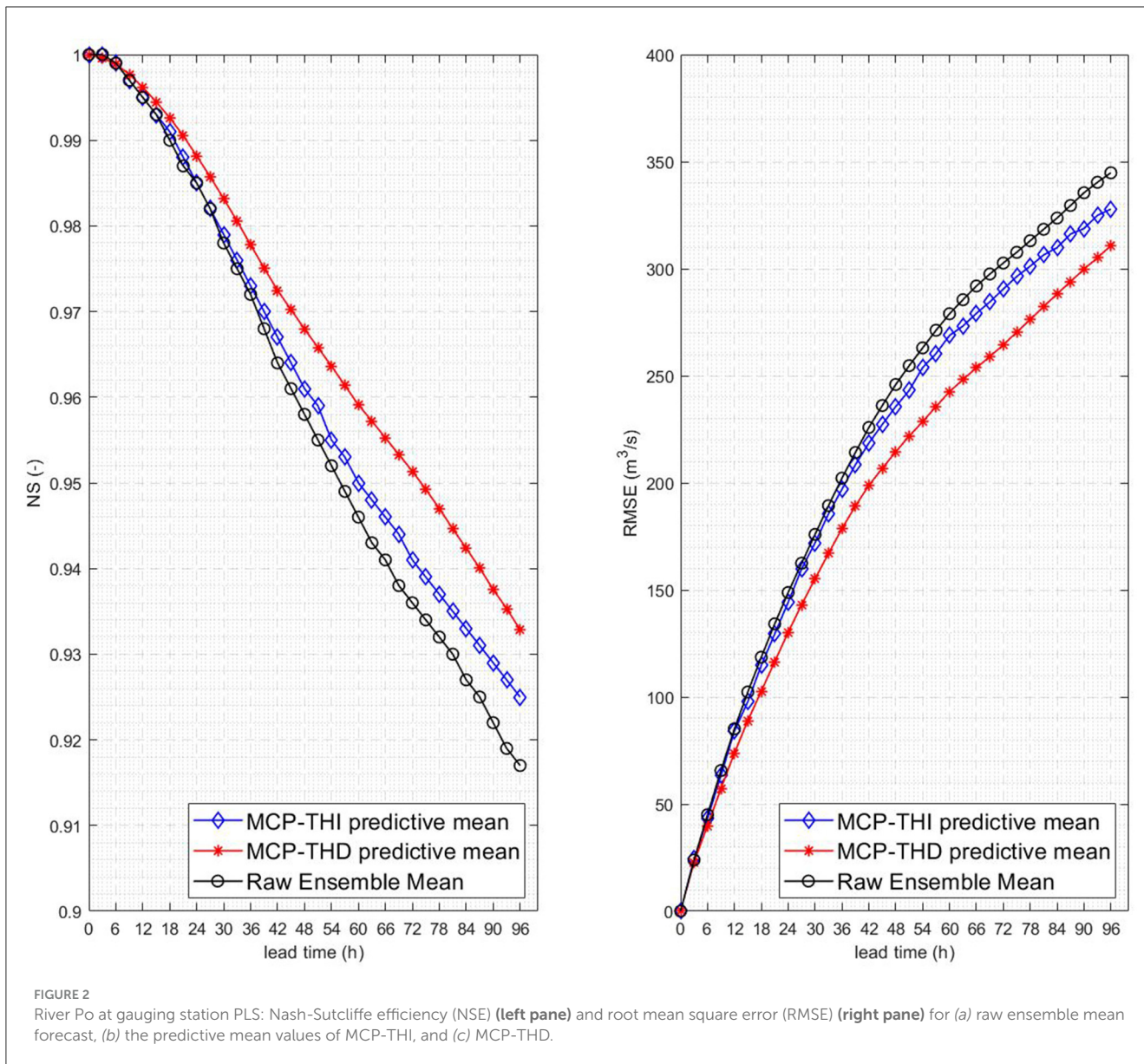
Table 2 compares the values of the performance indicators for MCP-THI and MCP-THD at different probability of threshold exceedance and three flood warning levels, respectively. In addition, here we emphasize the higher forecast skill of the time-horizon dependent approach, which is expressed in terms of larger Probability of Detection (POD), and smaller false alarm rates (FAR).

Figure 5 shows a forecast performance diagram. Performance diagrams were introduced by (Roebber, 2009) and provide a more concise presentation of the results. The diagram plots success ratio ( $SR = 1 - FAR$ ) on the abscissa and probability of detection (POD) on the ordinate axis, along with isolines representing BIAS and Threat Score (TS), which can be expressed as functions of these measures. This type of representation facilitates a clearer visual assessment of the forecast's quality and potential biases.

## 4 Discussion

Biondi and Todini (2018) compared the uni- and multivariate MCP of uncertainty, with conditional mean and variance specified by Equation (16), against raw ensemble mean and model averaging



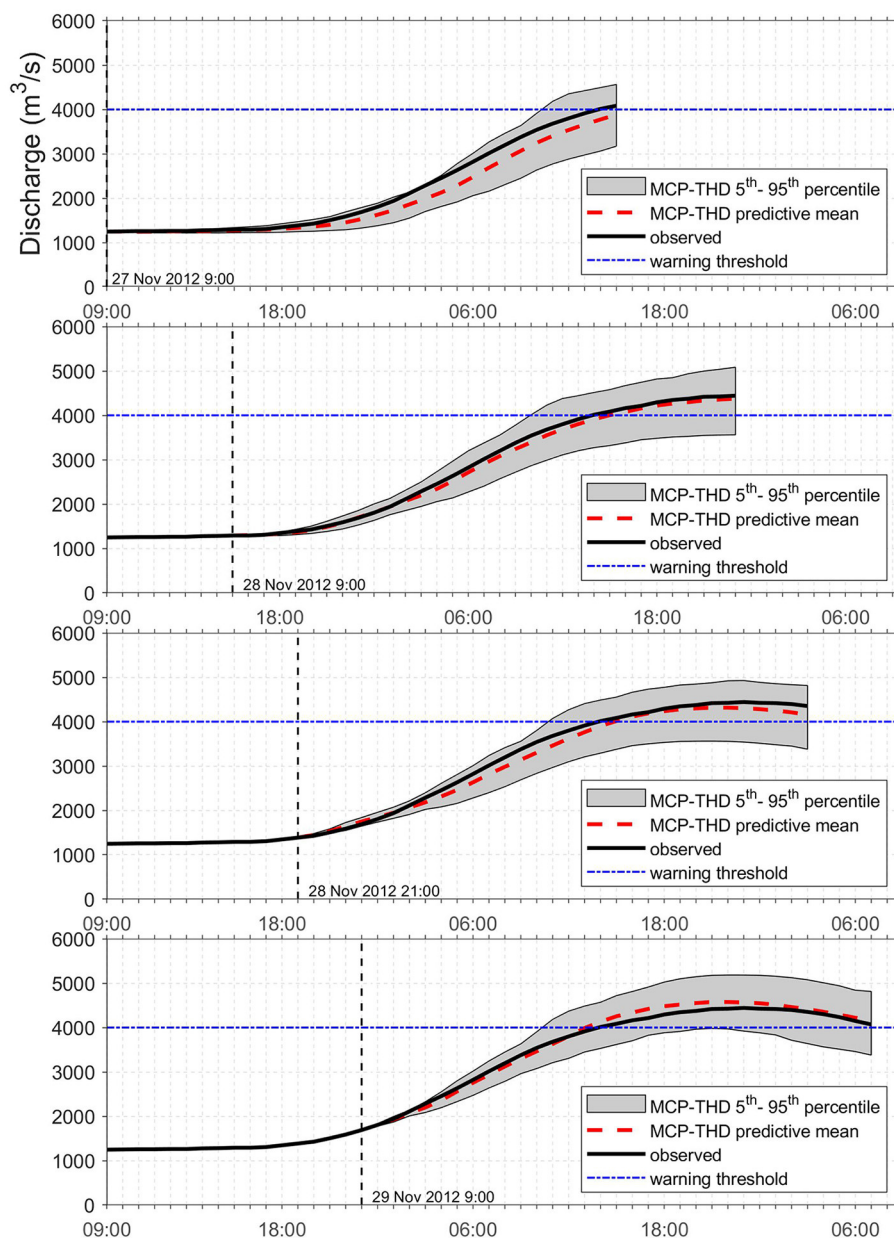


with uniform and non-uniform Bayesian weighting for time-horizon independent applications. Here, we extend the time horizon-independent multivariate MCP to allow for time-horizon dependency. While this concept was adopted earlier by [Barbetta et al. \(2017\)](#), we revisit the main statistical concepts and provide more discussion on underlying assumptions.

The relevance of time-horizon dependent forecasting lies in the informative added value provided to a forecaster by considering the temporal correlation of the flood peak propagation process at a site. Instead of outputting just the probability of threshold exceedance given a prediction, the user is enabled to evaluate how such likelihood evolves chronologically as the event approaches. Such insights provide additional support for planning mitigating measures under operational conditions.

The most important assumption of the MCP lies in the hypothesis of modeling the dependence structure among predictors and the predictand over different time horizons as multivariate

Gaussian copula, which is obtainable from standard normal images of the original marginal distributions of variates. Validating this hypothesis by measuring the distance between a multivariate Gaussian density model and the true multivariate distribution remains difficult in practice ([Herr and Krzysztofowicz, 2005](#)), mainly because of the dimensionality of the problem. However, the multivariate dependency assumption is attractive due to analytical tractability of conditional distributions. Its validity has been assessed indirectly by inter-comparison of performance indicators RMSE, MAE, and NSE for the time-horizon independent MCP against those of raw ensemble mean predictions and Bayesian model averaging ([Biondi and Todini, 2018](#)). Here, we show that the introduction of time-horizon dependency further increases processor performance as demonstrated through the indicators in [Table 1](#) and standard skill assessment criteria in [Table 2](#) and [Figure 5](#). In particular, the number of false and missed alarms of MCP-THD is smaller than for MCP-THI, leading to



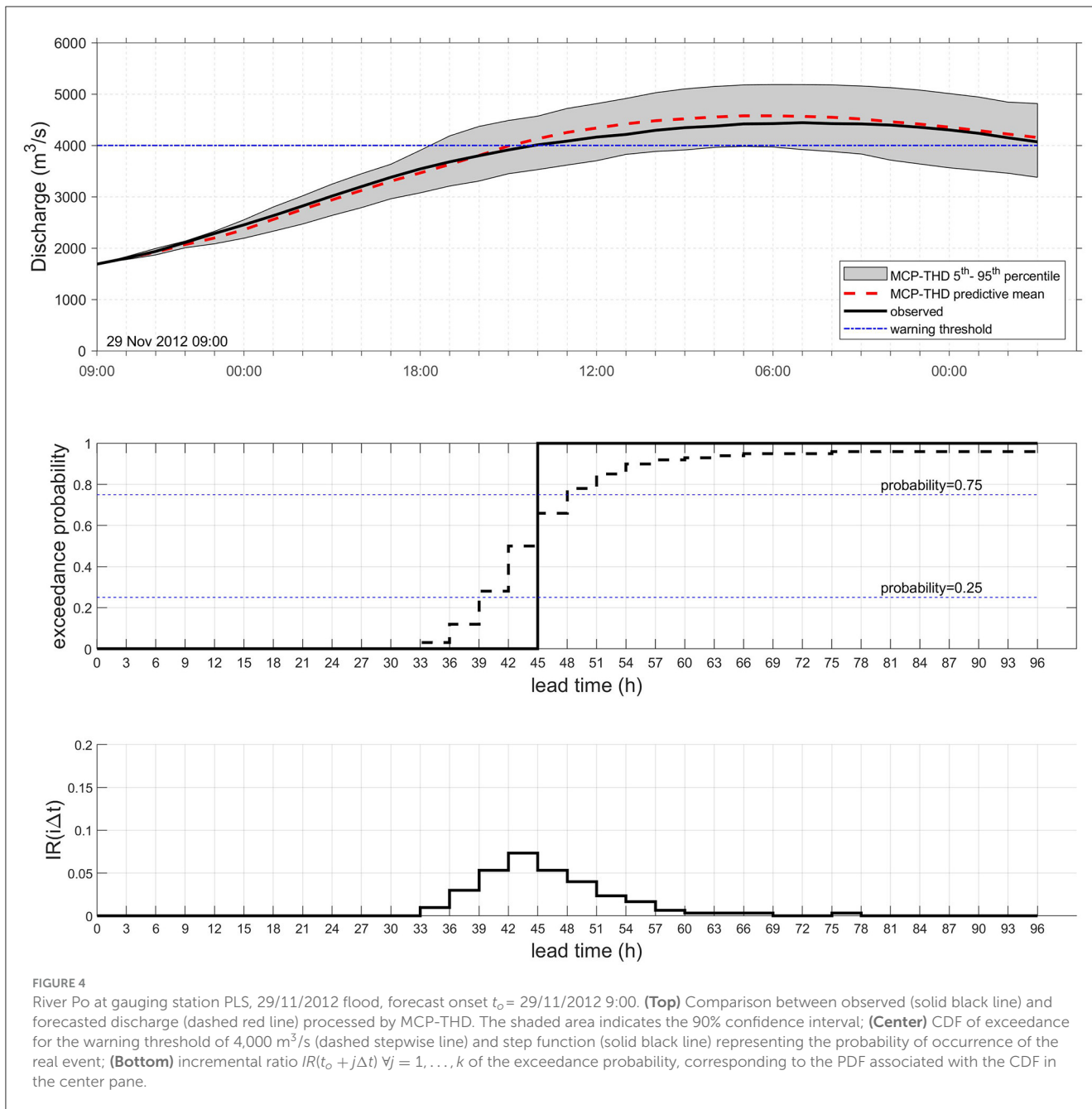
**FIGURE 3** River Po at gauging station PLS, 29/11/2012 flood, evolution of the forecast with onset times  $t_o = 27/11/2012$  21:00,  $28/11/2012$  09:00,  $28/11/2012$  21:00, and  $29/11/2012$  09:00. The shaded area indicate the 90% confidence interval, the solid black curve observations, and the dashed red curve the MCP-THD predictive mean. The 4,000 m<sup>3</sup>/s flood warning level is marked by the blue horizontal lines.

an overall higher threat score (TS) defined by Equation (A.2). The good performance of MCP-THD is corroborated by the concentration of POD vs. 1-FAR combinations for MC-THD sited in the upper right corner of the performance diagram in Figure 5.

In the same context, we also mention that the chosen test bed system refers to a large Italian river, which is characterized by a predominantly homoscedastic (constant variance) dependency relationship among predictors and the predictand. Smaller systems, however, tend to show a behavior in which the variance changes across the variable range, and needs to be correspondingly modeled through the chosen dependency structure (Coccia and Todini,

2011). An further elaboration on this topic for the MCP-THD, however, lies outside the scope of the present research.

While probabilities of event occurrence do not depend on the probability space in which the variables are represented, an inverse transformation of the predictive Gaussian distribution  $\phi(\eta|\hat{\eta})$  into its image  $f(\mathbf{h}|\hat{\mathbf{h}})$  in the space of origin is important in an operational setting, because forecasters prefer to work with quantities and scales they are acquainted with. Although not necessary to assess the probability of occurrence of a threshold exceedance in the forecasting horizon, which can be done in the Gaussian space, where the multivariate joint distribution is known, a multivariate probability density function in the space of origin would be



obtainable through Gibbs sampling (Geman and Geman, 1984). Nevertheless, such density may not be required in practice; whereas the multivariate analysis of exceedances can be performed in the Gaussian space, the marginal conditional distribution shown in Figure 4 yields operationally fully useable probabilistic forecasts of flow or stage.

The property of self-calibration for the MCP is assured in virtue of its definition through a joint density  $\phi(\eta, \hat{\eta})$ . Marginalization of the joint density with respect to predictors leads naturally to the density of historic streamflow observations, thus assuring that an MCP always falls back onto the unconditional long-term distributions of observation in the absence of informative predictors.

## 5 Conclusions

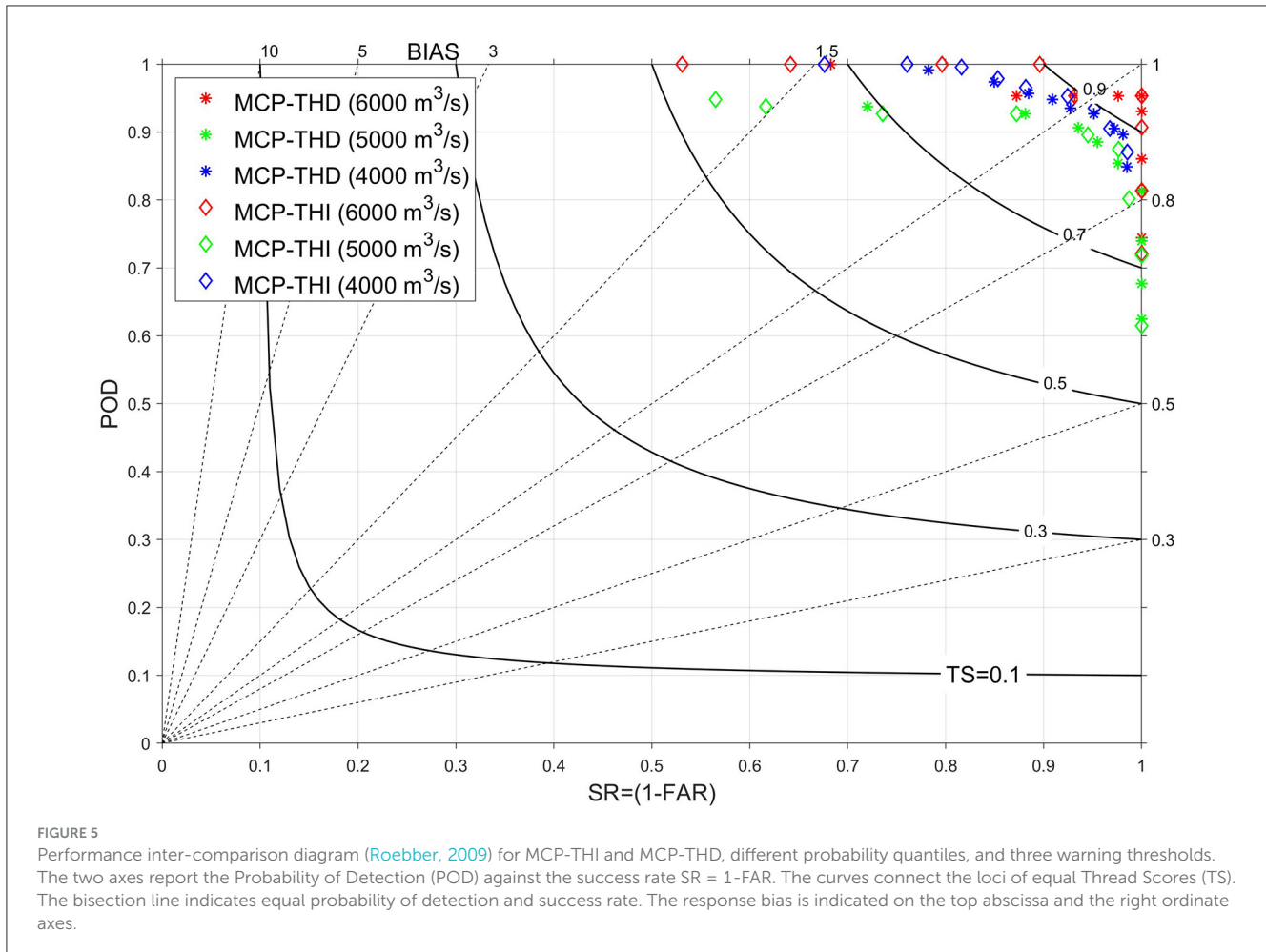
The study describes and promotes the use of a statistical processor designed to estimate the predictive uncertainty for an ensemble of stream flow predictors over a series of consecutive time steps across the forecasting horizon. We have demonstrated that such an extension of the original MCP to a multi-model and now time-horizon dependent version is straightforward and allows one to deal with high dimensionality problems. In the case study presented in this study, the dimension is 17 (1 series of observations plus 16 model ensemble members) times 32 (96 h forecasting horizon using 3 h time steps). In addition to the cross-correlations among multiple predictors, the

TABLE 2 Contingency table showing the forecasting skill for three warning levels. The results refer to different exceedance probability quantiles for MCP-THI and MCP-THD.

$P(h > h^*)$	Hits		False alarms		Missed alarms		Threat score		POD		FAR	
$h^* = 4,000 \text{ m}^3/\text{s}$	THI	THD	THI	THD	THI	THD	THI	THD	THI	THD	THI	THD
0.1	232	230	111	64	0	2	0.676	0.777	1.000	0.991	0.324	0.218
0.2	232	226	73	40	0	6	0.761	0.831	1.000	0.974	0.239	0.150
0.3	231	222	52	29	1	10	0.813	0.851	0.996	0.957	0.184	0.116
0.4	227	220	39	22	5	12	0.838	0.866	0.978	0.948	0.147	0.091
0.5	224	217	30	17	8	15	0.855	0.871	0.966	0.935	0.118	0.073
0.6	221	215	18	11	11	17	0.884	0.885	0.953	0.927	0.075	0.049
0.7	212	210	13	6	20	22	0.865	0.882	0.914	0.905	0.058	0.028
0.8	210	208	7	4	22	24	0.879	0.881	0.905	0.897	0.032	0.019
0.9	202	197	3	3	30	35	0.860	0.838	0.871	0.849	0.015	0.015
$h^* = 5,000 \text{ m}^3/\text{s}$	Hits		False alarms		Misses		Threat score		POD		FAR	
0.1	91	90	70	35	5	6	0.548	0.687	0.948	0.938	0.435	0.280
0.2	90	89	56	12	6	7	0.592	0.824	0.938	0.927	0.384	0.119
0.3	89	87	32	6	7	9	0.695	0.853	0.927	0.906	0.264	0.065
0.4	89	85	13	4	7	11	0.817	0.850	0.927	0.885	0.127	0.045
0.5	86	82	5	2	10	14	0.851	0.837	0.896	0.854	0.055	0.024
0.6	80	78	3	0	16	18	0.808	0.813	0.833	0.813	0.036	0.000
0.7	77	71	1	0	19	25	0.794	0.740	0.802	0.740	0.013	0.000
0.8	69	65	0	0	27	31	0.719	0.677	0.719	0.677	0.000	0.000
0.9	59	60	0	0	37	36	0.615	0.625	0.615	0.625	0.000	0.000
$h^* = 6,000 \text{ m}^3/\text{s}$	Hits		False alarms		Misses		Threat score		POD		FAR	
0.1	43	43	38	20	0	0	0.531	0.683	1.000	1.000	0.469	0.317
0.2	43	41	24	6	0	2	0.642	0.837	1.000	0.953	0.358	0.128
0.3	43	41	11	3	0	2	0.796	0.891	1.000	0.953	0.204	0.068
0.4	43	41	5	1	0	2	0.896	0.932	1.000	0.953	0.104	0.024
0.5	41	41	3	0	2	2	0.891	0.953	0.953	0.953	0.068	0.000
0.6	41	40	0	0	2	3	0.953	0.930	0.953	0.930	0.000	0.000
0.7	39	37	0	0	4	6	0.907	0.860	0.907	0.860	0.000	0.000
0.8	35	35	0	0	8	8	0.814	0.814	0.814	0.814	0.000	0.000
0.9	31	32	0	0	12	11	0.721	0.744	0.721	0.744	0.000	0.000

processor also exploits the temporal correlations across successive forecasting time steps. Acknowledging such correlation has proven to yield multiple advantages with respect to conventional time-horizon independent processing. These can be summarized as follows:

- Time-horizon dependent forecasting does not only provide the probability of threshold exceedance for a predictand at a set future point in time but also allows tracking the flood wave across time in probabilistic terms.
- By additionally acknowledging the autocorrelation structure of the flood process, which is especially strong in highly predictable river system with slow flood motion, the forecasting error, measured in terms of RMSE and NSE, can be reduced considerably over the forecasting horizon in comparison to time-independent processing.
- Similarly, standard skill measurement indicators, including the False Alarm Rate (FAR) and the Probability of Detection (POD), show improvements over time-horizon independent processing for the selected river system.



- The advantages outlined above may diminish when smaller river systems with fast flood propagation behavior are considered. In this case, the dependency structure between predictors and the predictand becomes heteroscedastic and characterized by frequent outliers, two properties which lead to increases of the prediction error.
- To adapt the THD processor to applications with heteroscedastic behavior, the dependency needs to be modeled over different sub-ranges of the joint distributions. While this issue has been addressed preliminarily by Coccia and Todini (2011) by means of truncation, further research is needed.

Investigation, Methodology, Software, Writing—original draft, Writing—review & editing. ET: Conceptualization, Writing—original draft, Writing—review & editing.

### Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The work has been partially funded by Deutsche Forschungsgemeinschaft grant RE 3834/10-1 awarded to PR.

### Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

### Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Author contributions

PR: Conceptualization, Writing—original draft, Writing—review & editing. DB: Conceptualization, Data curation,

### Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Alpert, M., and Raiffa, H. (1982). "A progress report on the training of probability assessors," in *Judgment under Uncertainty: Heuristics and Biases*, eds. D. Kahneman, P. Slovic, and A. Tversky (Cambridge University Press), 294–305.
- Barbetta, S., Coccia, G., Moramarco, T., Brocca, L., and Todini, E. (2017). The multi temporal/multi-model approach to predictive uncertainty assessment in real-time flood forecasting. *J. Hydrol.* 551, 555–576. doi: 10.1016/j.jhydrol.2017.06.030
- Biondi, D., and Todini, E. (2018). Comparing hydrological postprocessors including ensemble predictions into full predictive probability distribution of stream flow. *Water Resour. Res.* 54, 9860–9882. doi: 10.1029/2017WR022432
- Brown, J. D., and Seo, D.-J. (2010). A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *J. Hydrometeorol.* 11, 642–665. doi: 10.1175/2009JHM1188.1
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., et al. (2008). Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. *Adv. Water Resour.* 31, 1309–1324. doi: 10.1016/j.advwatres.2008.06.005
- Coccia, G., and Todini, E. (2011). Recent developments in predictive uncertainty assessment based on the Model Conditional Processor approach. *Hydrol. Earth Syst. Sci.* 15, 3253–3274. doi: 10.5194/hess-15-3253-2011
- Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pat. Anal. Machine Intell.* 6, 721–741. doi: 10.1109/TPAMI.1984.4767596
- Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Rev.* 133, 1098–1118. doi: 10.1175/MWR2904.1
- Hamill, T. M., and Whitaker, J. S. (2006). Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Mon. Wea. Rev.* 134, 3209–3229. doi: 10.1175/MWR3237.1
- Hemri, S., Fundel, F., and Zappa, M. (2013). Simultaneous calibration of ensemble river flow predictions over an entire range of lead times. *Water Resour. Res.* 49, 6744–6755. doi: 10.1002/wrcr.20542
- Hemri, S., and Klein, B. (2017). Analog-based postprocessing of navigation-related hydrological ensemble forecasts. *Water Resour. Res.* 53, 9059–9077. doi: 10.1002/2017WR020684
- Hemri, S., Lisniak, D., and Klein, B. (2015). Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resour. Res.* 51, 7436–7451. doi: 10.1002/2014WR016473
- Herr, H. D., and Krzysztofowicz, R. (2005). Generic probability distribution of rainfall in space: the bivariate model. *J. Hydrol.* 306, 234–263. doi: 10.1016/j.jhydrol.2004.09.011
- Herr, H. D., and Krzysztofowicz, R. (2010). Bayesian ensemble forecast of river stages and ensemble size requirements. *J. Hydrol.* 387, 151–164. doi: 10.1016/j.jhydrol.2010.02.024
- Herr, H. D., and Krzysztofowicz, R. (2015). Ensemble bayesian forecasting system part I: theory and algorithms. *J. Hydrol.* 524, 789–802. doi: 10.1016/j.jhydrol.2014.11.072
- Hersbach, R. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast.* 15, 559–570. doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2
- Kalnay, E., Li, H., Miyoshi, T., Yang, S. C., and Ballabrera-Poy, J. (2007). 4-D-Var or ensemble Kalman filter? *Tellus A* 59, 758–773. doi: 10.1111/j.1600-0870.2007.00261.x
- Kelly, K. S., and Krzysztofowicz, R. (2000). Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* 36, 2643–2653. doi: 10.1029/2000WR900061
- Kitanidis, P. K., and Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. *Water Resour. Res.* 16, 1025–1033.
- Koenker, R. (2005). *Quantile Regression*. New York, NY: Econometric Society Monographs, Cambridge University Press.
- Krzysztofowicz, R. (1999). Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* 35, 1757–1760.
- Krzysztofowicz, R. (2001a). The case for probabilistic forecasting in hydrology. *J. Hydrol.* 249, 2–9. doi: 10.1016/S0022-1694(01)00420-6
- Krzysztofowicz, R. (2001b). Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. *J. Hydrol.* 249, 69–85. doi: 10.1016/S0022-1694(01)00413-9
- Krzysztofowicz, R. (2014). Probabilistic flood forecast: exact and approximate predictive distributions. *J. Hydrol.* 517, 643–651. doi: 10.1016/j.jhydrol.2014.04.050
- Krzysztofowicz, R., and Kelly, K. S. (2000). Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* 36, 3265–3277. doi: 10.1029/2000WR900108
- Liu, Z., Martina, M. L. V., and Todini, E. (2005). Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian catchment. *Hydrol. Earth Syst. Sci.* 9, 347–364. doi: 10.5194/hess-9-347-2005
- Maranzano, C. J., and Krzysztofowicz, R. (2004). Identification of likelihood and prior dependence structures for hydrologic uncertainty processor. *J. Hydrol.* 290, 1–21. doi: 10.1016/j.jhydrol.2003.11.021
- Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T. (1996). The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Royal Meteorol. Soc.* 122, 73–119.
- Pagano, T. C., Shrestha, D. L., Wang, Q. J., Robertson, D., and Hapuarachchi, P. (2013). Ensemble dressing for hydrological applications. *Hydrol. Process.* 27, 106–116. doi: 10.1002/hyp.9131
- Pagano, T. C., Wood, A. W., Ramos, M.-H., Cloke, H. L., Pappenberger, F., Clark, M. P., et al. (2014). Challenges of operational river forecasting. *J. Hydrometeorol.* 15, 1692–1707. doi: 10.1175/JHM-D-13-0188.1
- Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian Model Averaging to calibrate forecast ensembles. *Mon. Wea. Rev.* 133, 1155–1174. doi: 10.1175/MWR2906.1
- Reggiani, P., and Boyko, O. (2019). A Bayesian processor of uncertainty for precipitation forecasting using multiple predictors and censoring. *Mon. Wea. Rev.* 147, 1–17. doi: 10.1175/MWR-D-19-0066.1
- Reggiani, P., Renner, M., Weerts, A. H., and van Gelder, P. A. H. J. M. (2009). Uncertainty assessment via bayesian revision of ensemble streamflow predictions in the operational river rhine forecasting system. *Water Resour. Res.* 45:WR006758. doi: 10.1029/2007WR006758
- Reggiani, P., and Weerts, A. H. (2008). Probabilistic quantitative precipitation forecast for flood prediction: an application. *J. Hydrometeorol.* 3, 446–462. doi: 10.1175/2007JHM858.1
- Roeber, P. J. (2009). Visualizing multiple measures of forecast quality. *Wea. Forecast.* 24, 601–608. doi: 10.1175/2008WAF2222159.1
- Schefzik, R., Thorarindottir, T. L., and Gneiting, T. (2013). Uncertainty quantification in complex simulation models using Ensemble Copula Coupling. *Stat. Sci.* 28, 616–640. doi: 10.1214/13-STS443
- Seo, D.-J., Koren, V., and Cajina, N. (2003). Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *J. Hydrometeorol.* 4, 627–641. doi: 10.1175/1525-7541(2003)004<0627:RVAOHA>2.0.CO;2
- Stelling, G. S., and Verwey, A. (2006). *Numerical Flood Simulation, Chapter 16*. Hoboken, NJ: John Wiley and Sons, Ltd.
- Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manag.* 36, 3265–3277. doi: 10.1080/15715124.2008.9635342
- Todini, E. (2012). From HUP to MCP: analogies and extended performances. *J. Hydrol.* 477, 32–43. doi: 10.1016/j.jhydrol.2012.10.037
- Toth, Z., and Kalnay, E. (1993). Ensemble forecasting at NMC: the generation of perturbations. *Bullet. Am. Meteorol. Soc.* 74, 2317–2330.
- Verkade, J., Brown, J., Reggiani, P., and Weerts, A. (2013). Post-processing ecwf precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol.* 501, 73–91. doi: 10.1016/j.jhydrol.2013.07.039
- Verkade, J. S., Brown, J. D., Davids, F., Reggiani, P., and Weerts, A. H. (2017). Estimating predictive hydrological uncertainty by dressing deterministic and ensemble forecasts; a comparison, with application to meuse and rhine. *J. Hydrol.* 555, 257–277. doi: 10.1016/j.jhydrol.2017.10.024
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. *Water Resour. Res.* 41:WR003059. doi: 10.1029/2004WR003059
- Wilks, D. S. (1995). *Statistical Methods in the Atmospheric Sciences: An Introduction*. Cambridge, MA: Academic Press.
- Wood, A. W., and Schaake, J. C. (2008). Correcting errors in streamflow forecast ensemble mean and spread. *JHM* 9, 132–148. doi: 10.1175/2007JHM862.1

## Appendix A: performance indicators

TABLE A1 Contingency table summarizing possible outcomes of a forecast.

Observed	True	False
Predicted		
True	Correct alarm (CA)	False alarm (FA)
False	Missed alarm (MA)	Correct non-alarm (CN)

The deterministic performance indicators used in Section (3.3) for assessing the skill of the probabilistic forecasts are defined as follows.

$$\text{Probability of Detection (Hit Rate): } POD = \frac{CA}{CA + MA} \quad (A.1)$$

where *CA* (see Table A1) is the total number of correct alarms and *MA* are the missed alarms. The threat score is given by

the ratio:

$$\text{Threat score (Critical Success Ratio): } TS = \frac{CA}{CA + MA + FA} \quad (A.2)$$

where *FA* is the total number of false alarms. The False Alarm Rate is:

$$\text{False Alarm Rate: } FAR = \frac{FA}{CA + FA} \quad (A.3)$$

while the False Positive Rate is:

$$\text{False Positive Rate: } FPR = \frac{FA}{FA + CN} \quad (A.4)$$

with *CN* the number of correct negative forecasts. Finally the response bias is defined as:

$$\text{Bias: } BIAS = \frac{FA + CA}{CA + MA} \quad (A.5)$$

All indicators are evaluated over the sample of forecasts instances.