


A parsimonious post-processor for uncertainty evaluation of ensemble precipitation forecasts: an application to quantitative precipitation forecasts for civil protection purposes

Daniela Biondi ^{a,*}, Ezio Todini^b and Angela Corina^c

^a Department of Informatics, Modelling, Electronics and Systems Engineering, University of Calabria, Ponte P. Bucci Cubo 41/b, 87036 Rende, CS, Italy

^b Italian Hydrological Society, Bologna, Italy

^c Italian Civil Protection Department, Roma, Italy

*Corresponding author. E-mail: daniela.biondi@unical.it

 DB, 0000-0002-3570-9605

ABSTRACT

In this study, we present a post-processing technique designed to assess conditional predictive uncertainty for quantitative precipitation forecasts (QPFs) that extends to the multivariate predictor case. The theoretical foundation of the mixed binary-continuous precipitation process representation for a single predictor is coupled with the univariate version of the model conditional processor (MCP) that allows considering ensemble forecasts while maintaining a parsimonious approach. The experiment set-up is based on a weather vigilance zone (WVZ) of the national warning system in southern Italy and the relative QPFs issued by the Italian Civil Protection Department. Various aspects of the quality of the probabilistic forecast from the uncertainty processor were evaluated. The results indicate that post-processed forecasts can provide improved performances in terms of accuracy and reliability, tend to correct bias and are generally less under-dispersive than raw forecasts for the investigated area. Furthermore, we explored the possibility of issuing warnings employing the full predictive distribution and moving to the use of probabilistic thresholds here identified through a receiver operating characteristic (ROC) analysis. Considering the probability of exceeding a critical rainfall value allowed successful discrimination between events and non-events for critical precipitation occurrences and proved to be a valuable approach to decision-makers and information providers.

Key words: multivariate analysis, predictive uncertainty, probabilistic thresholds, rainfall statistical post-processor

HIGHLIGHTS

- A post-processor to assess predictive uncertainty that extends to the multivariate rainfall predictors case is presented.
- The mixed binary-continuous precipitation process is coupled with the UMCP that allows considering ensemble forecasts.
- Post-processed forecasts improved accuracy and reliability.
- Moving to probabilistic forecasts and probabilistic thresholds enhanced skill statistics to correctly forecast critical events.

1. INTRODUCTION

The process of providing accurate and timely forecast of impending events in a systematic way is crucial for an efficient and effective implementation of early warning systems (EWSs) aimed at reducing the impact of natural hazards like floods and landslides (UNPD 2018). The raw forecasts are often single deterministic or ensemble forecasts that lack provision of a complete characterization of the uncertainty needed by rational decision-makers and by information providers (Demeritt *et al.* 2007; Cloke & Pappenberger 2009). Statistical post-processors have been proposed to remove systematic errors as well as to improve the description of the predictive probability distribution, often by transforming a non-probabilistic forecast into a probabilistic one and increasing forecast reliability. Both goals are in many cases pursued by assessing the predictive uncertainty defined as the probability distribution of the true future state of the predictand conditional on appropriate predictors (Krzysztofowicz 1999). The value added by statistical post-processing of forecasts has been demonstrated over a variety of forecasting applications ranging from river stages and floods (e.g., Krzysztofowicz 2002; Todini 2008; Reggiani & Weerts 2009; Madadgar *et al.* 2014; Hemri *et al.* 2015) to weather-climate data (Hamill *et al.* 2004; Wilks & Hamill 2007; Verkade *et al.* 2013; Reggiani

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

et al. 2021). It is worth noting that post-processing meteorological forecasts imply dealing with the inherent uncertainty of numerical weather prediction (NWP) models, given the imperfect knowledge of initial conditions and the nonlinearity of the processes, as well as to account for the heteroscedasticity of forecast error and the discrepancy between spatial and temporal scales (Reggiani & Byoko 2019). Li *et al.* (2017) reviewed a variety of post-processing approaches in meteorological forecasting that have been proposed and tested over the past years, as follows:

- empirical methods such as quantile mapping (QM) (Panofsky & Brier 1968; Piani *et al.* 2010) and analogue method (Hamill & Whitaker 2006), also corresponding to the early approaches, that can be largely assumed as data-driven thus providing no process insights (Koutsoyiannis *et al.* 2008);
- regression-based approaches like logistic regression (Wilks 2009), quantile regression (Bremnes 2004; Weerts *et al.* 2011) and model output statistics (MOS) (Glahn & Lowry 1972; Wilks 2005);
- methodologies that provide the predictive probability of the variable of interest conditional on a single forecast, e.g., the Bayesian processor of output (Krzysztofowicz & Maranzano 2006), the ensemble preprocessor (EPP, Schaake *et al.* 2007), the Bayesian joint probability (Wang *et al.* 2009; Robertson *et al.* 2013);
- multivariate post-processing methods that combine outputs from multiple models or ensemble weather forecasts to achieve better forecast skill like, for example, the Bayesian processor of ensemble (Krzysztofowicz 2008), the ensemble model output statistics (EMOS; Gneiting *et al.* 2005), the Bayesian model averaging (BMA; Raftery *et al.* 2005; Sloughter *et al.* 2007), and the model conditional processor (Todini 2008; Reggiani & Byoko 2019).

Post-processors specifically aiming at providing a probabilistic quantitative precipitation forecast (PQPF) pose an extra challenge in comparison to other hydrometeorological processes given the intermittent character of the variable at short time scales. The analysis of intermittency and the modelling of the occurrence process of rainfall have attracted the interest of many studies; most of the post-processors developed for rainfall forecasts employ a widely used representation for the precipitation process as a mixed discrete/continuous distribution (e.g., Bell 1987; Koutsoyiannis 2006; Bárdossy & Pegram 2009; Serinaldi 2009; Tsoukalas *et al.* 2019).

Describing the precipitation process as a mixture of binary and continuous random variables rests on a distribution consisting of two components: the probability of a dichotomous variable such as precipitation occurrence (PoP) and the distribution of the continuous process representing the amount of precipitation (DoA). If the random variable X describes the precipitation depth, with realization x , the mixed binary-continuous process assigns a probability mass of $(1 - \text{PoP})$ to the event $X = 0$ and spreads the remaining probability mass, i.e., the probability of precipitation occurrence $\text{PoP} = P(X > 0)$, over the complement of its support, that is the interval $(0, \infty)$ for the precipitation process (Kelly & Krzysztofowicz 2000). Usually, precipitation event occurrence is estimated as climatological relative frequency from a sample of observations, while the continuous process is modelled with a suitable parametric distribution function. Exponential, Gamma and Weibull distributions are applied in the vast majority of the existing precipitation models; nevertheless, with the focus on the statistical features of the extremes of rainfall process at small time scale (e.g., daily or hourly), many studies based on very long time series of observations have concluded that light-tailed distributions often lead to an underestimation of high precipitation amounts and that heavy-tailed distributions (e.g., Pareto type distributions, generalized beta of the second kind – GB2) are generally in better agreement with observed precipitation extremes (Mielke & Johnson 1974; Todini & Di Bacco 1997; Evin *et al.* 2018; Koutsoyiannis 2021).

Recently, Reggiani & Byoko (2019) extended the use of MCP to multivariate mixed binary-continuous distribution variables, specifically daily precipitation, by transforming the intermittent random process into a continuous one: the intermittent multivariate precipitation process is considered as censored and turned into a continuous one by replacing censored data via imputation using a Gibbs sampler. The sample is augmented into the negative range, recovering the true, a priori unknown variance-covariance structure of the full uncensored sample. The MCP was also applied to the precipitation process in Massari *et al.* (2019) to provide its probability distribution conditional on satellite precipitation estimates and a soil moisture-based rainfall product. Nevertheless, the MCP was applied here only to cases when both predictand and predictors are non-zero during calibration and when predictors are non-zero during validation. To account for the intermittency issue, in addition to MCP, a categorical case integration method based on the detection capability of the predictors was considered.

In the present study, the univariate version of the MCP (UMCP), introduced in Biondi & Todini (2018) for probabilistic flow forecasting, is adapted to post-processing of ensemble rainfall forecasts. Here, we coupled the theoretical foundation of the mixed binary-continuous precipitation process representation for a single predictor (Herr & Krzysztofowicz 2005) with the UMCP that allows ensemble forecasts to be considered. By summarizing the ensemble forecast with its mean, the post-processor provides the

predictive uncertainty about a predictand given ensemble forecasts in a univariate model framework and, at the same time, properly takes into account the information from the ensemble spread. The principal advantage of the proposed approach is that the extension of the single predictor case to multivariate does not entail the increasing of the number of mutually conditional PoP combinations and conditional/marginal distributions. Moreover, the method allows inclusion of the nonstationary behaviour of the precipitation process, the non-Gaussian form of the probability distributions of the random variables, the nonlinear and heteroscedastic dependence structure between the forecast and the predictand and does not require invoking the exchangeability of the ensemble members as other post-processors. A weather vigilance zone (WVZ) of the Italian national alert system coordinated by the Civil Protection Department was considered as a case study to test the performance of the approach as well as potential implications for operational forecasting of critical rainfall that may trigger floods, landslides or other damaging events. The post-processor is applied to rainfall forecasts issued daily by the department for civil protection purposes in the form of a range of quantitative precipitation forecasts (QPFs). In the current operational setting, the forecasts are subjective in that they are made judgmentally by expert forecasters based on information from multiple sources, including output from numerical weather prediction (NWP) models. The aim of this paper is the assessment of the ability of UMCP to be integrated with mixed binary-continuous distribution models in order to obtain a parameter-parsimonious and computationally efficient processor applicable in operational settings and to provide a PQPF that can be useful to decision-makers and information providers. Furthermore, we explored the possibility of issuing warnings by comparing resulting PQPF to a critical probability threshold value identified through a receiver operating characteristic (ROC) analysis.

The paper is organized as follows. Section 2 presents the case study and the available data. Section 3 details the assumptions for the mixed binary-continuous precipitation process representation and the univariate version of the MCP. Results are presented in Section 4. Section 5 summarizes implications and potential advantages of the proposed approach.

2. STUDY AREA AND DATA

The Civil Protection Department coordinates the Italian national warning system that involves the regions through a network of functional centers for collecting and sharing of meteorological, hydrogeological and hydraulic data, but also for real-time analysis and monitoring, impact-based forecasting and dissemination of warning messages. The forecasting phase mainly relies on the combined use of rainfall thresholds and QPFs. The National Weather Vigilance Bulletin, issued by the department, is an informative tool for the civil defence national system that highlights relevant adverse weather phenomena expected up to the next 24 hours, and the tendency for the following days. In the operational setting for impact assessment, the department issues forecasts in the form of a range of QPFs over WVZs covering the national territory.

WVZs represent wide areas according to geographic, physical and climatic boundaries often going beyond regional limits. An area of about 5,000 km² located in central-north-eastern Calabria (south Italy) corresponding to WVZ #36 (numbering valid until December 2016) and depicted in [Figure 1](#), was selected as case study.

Issued QPFs are a quantitative estimate of precipitation, elaborated judgmentally by expert forecasters based on information from multiple sources, including output from numerical weather prediction models. QPFs represent the best possible forecast information at the state in which they are issued and constitute the main technical-scientific contribution to allow daily critical assessments at the regional level, particularly where local meteorological services are not available. The reference meteorological model is the deterministic model developed by the ECMWF in the 00 UTC and/or 12 UTC run. The forecasters also evaluate the output from the high-resolution models of the COSMO operational chain of ARPA ER and other available global and local models. This study considers QPFs elaborated daily by the Italian Civil Protection Department for the period 2007–2015, that refer to the maximum value of cumulated point rainfall expected over the investigated area for different durations, namely, 3, 6 and 12 hours. For a given duration, the limits of a range, corresponding to given probabilities of exceedance, for the maximum cumulated rainfall over the WVZ are provided. Probability of exceedance values are available until November 2013. The forecasts can refer to phenomena expected up to midnight of the day of issuance or for the following day; also, the tendency of the day after can be provided. In this study, only forecasts related to the day of issuance are considered.

Furthermore, it is worth noting that:

- weather forecasts for civil protection purposes are specifically devoted to highlight potentially harmful situations to people or things; they are often the result of subjective evaluations and do not strictly correspond to the outputs of NWP models;

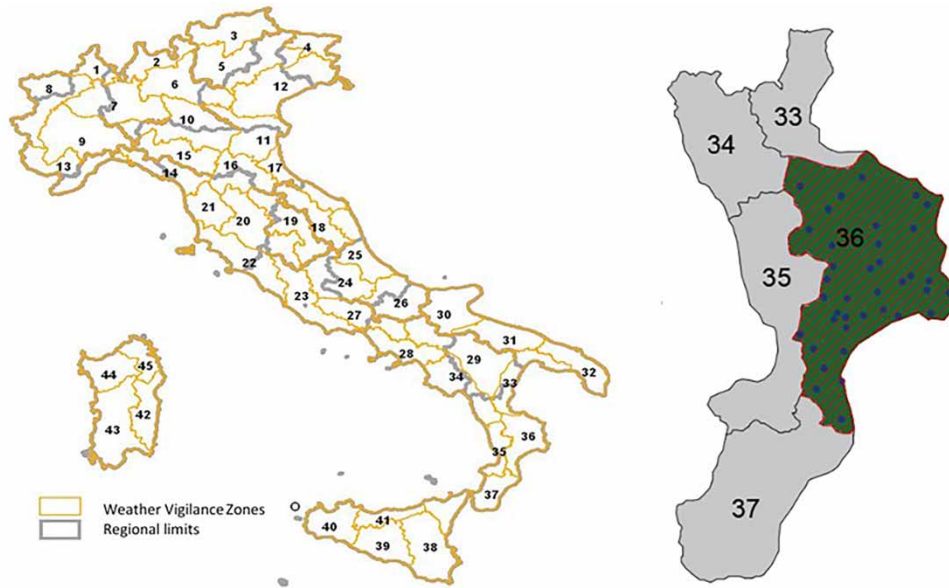


Figure 1 | Weather vigilance zone (WVZ) configuration valid until December 2016 and location of the investigated WVZ #36. Blue dots correspond to rain gauges. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/nh.2021.045>.

- time series of QPFs for the various precipitation durations are characterized by different sample size as daily bulletins do not always include evaluations for every duration;
- the lack of continuity in QPF time series for the different durations may not guarantee the presence of valuable data and high precipitation values could be missed (particularly relevant in the case of extreme events);
- available QPFs assume integer strictly positive values with an increment Δ of 5 mm.

The local warning systems are mostly based on the comparison of the QPFs with a set of rainfall thresholds that identifies precipitation critical values concerning landslides and/or flood occurrence and supports emission of warnings for the relative alert zone at three severity levels (ordinary, moderate and high).

Observed rainfall at 20 min time resolution, for the same period (2007–2015), is drawn by the rain gauge network covering the WVZ consisting of 39 stations (data available at <https://www.cfd.calabria.it/>). To compare coarser operational QPFs with observed data, information from each rain gauge has been used to obtain daily maximum cumulated rainfall values for different durations (specifically 3, 6 and 12 hours corresponding to durations of available QPFs) over the investigated WVZ.

Table 1 summarizes seasonality and sample size of QPFs and observed rainfall datasets. In Table 1, the number of data corresponding to observed maximum data is equal for all investigated durations and reveals the availability of a longer climatic sample of the predictand and a shorter sample of operational forecasts.

3. METHODS

3.1. The mixed binary-continuous precipitation process

Herr & Krzysztofowicz (2005) derived a generic bivariate probability distribution developed for forecasting precipitation in space with a single predictor. The bivariate distribution function is of mixed discrete-continuous form, with the discrete

Table 1 | Sample size of available QPF issuances and observed daily maximum rainfall over the investigated WVZ

| Duration | | Season | | | | Total |
|----------|--------------|--------|--------|--------|--------|-------|
| | | Spring | Summer | Autumn | Winter | |
| All | Observations | 819 | 828 | 828 | 812 | 3,287 |
| 3 h | Forecasts | 27 | 41 | 54 | 25 | 147 |
| 6 h | | 29 | 22 | 34 | 21 | 106 |
| 12 h | | 22 | 18 | 95 | 110 | 245 |

component modelling the occurrence or non-occurrence of precipitation and the continuous component modelling the amount of precipitation given that some precipitation occurs. A closed form of the distribution function of the predictand conditional on the predictor was derived with the meta-Gaussian approach in terms of the mutually conditional PoP occurrence, four marginal conditional distributions and a dependence parameter. The extension of the bivariate case to the multivariate one, as required in the case of multiple or ensemble predictors, poses some problems because of the large number of involved conditional PoP combinations and conditional marginal distributions (Reggiani & Byoko 2019). Here, we propose the coupling of the bivariate conditional dependence structure with the UMCP (Biondi & Todini 2018) to be used for predictive probability estimation in the case of multiple forecasts, adapting the methodology to the specific operational constraints of the selected case study. UMCP was designed for the post-processing of ensemble forecasts by using the expected value of the ensemble as the predictor, thus allowing its integration within the general framework of the bivariate dependence case.

Nevertheless, given the peculiarity of the case study and the relative operational constraints, the approach presented here refers to the case of non-zero predictors; indeed, available forecasts are integer strictly positive values and daily bulletins do not always include evaluations for each duration, thus, missing values are difficult to interpret. Having more information, the same methodology allows treating the ensemble of precipitation forecasts considering also the zero case for predictors, by integrating the UMCP in the complete bivariate conditional model (assuming the mean value of the ensemble as predictor) described in Herr & Krzysztofowicz (2005).

Referring to the specific case study, let X and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$ be the amounts of observed precipitation and m ensemble predictions, respectively, with $X \geq 0$ and $Y_i > 0 \forall i = 1, \dots, m$ given that provided forecasts when issued are always greater than zero. Let \bar{Y} be the average value of the predictions.

The conditional cumulative distribution functions (CDF) of X given $X > 0, \bar{Y} > 0$ and of \bar{Y} given $X > 0$ and $\bar{Y} > 0$ are:

$$F_X(x) = F_X(x|X > 0, \bar{Y} > 0) \quad (1)$$

$$F_{\bar{Y}}(\bar{y}) = F_{\bar{Y}}(\bar{y}|X > 0, \bar{Y} > 0) \quad (2)$$

and the corresponding density functions are denoted by their respective lower case letters.

Moreover, let $H_{\bar{Y}}(\bar{y})$ be the conditional CDF defined as:

$$H_{\bar{Y}}(\bar{y}) = H_{\bar{Y}}(\bar{y}|X = 0, \bar{Y} > 0). \quad (3)$$

The probability of a non-null occurrence of observed precipitation given a non-null prediction is:

$$p_{11} = P(X > 0|\bar{Y} > 0) \quad (4)$$

while

$$p_{01} = 1 - p_{11} = P(X = 0|\bar{Y} > 0). \quad (5)$$

Consequently, the marginal CDFs of X and \bar{Y} in this application can be defined as:

$$G_X(x) = p_{01} + p_{11}F_X(x) \quad (6)$$

$$G_{\bar{Y}}(\bar{y}) = p_{01}H_{\bar{Y}}(\bar{y}) + p_{11}F_{\bar{Y}}(\bar{y}) \quad (7)$$

The predictive distribution we want to estimate is the conditional CDF of X given the mean ensemble value $\bar{Y} = \bar{y}$ for $\bar{y} > 0$, that is:

$$G_{X|\bar{Y}}(x|\bar{Y} = \bar{y}, \bar{Y} > 0) = \frac{p_{01}h_{\bar{Y}}(\bar{y}) + p_{11}f_{\bar{Y}}(\bar{y})H_{X|\bar{Y}}(x|\bar{y})}{p_{01}h_{\bar{Y}}(\bar{y}) + p_{11}f_{\bar{Y}}(\bar{y})} \quad (8)$$

where $H_{X|\bar{Y}}(x|\bar{y})$ is the conditional CDF defined as:

$$H_{X|\bar{Y}}(x|\bar{y}) = H_{X|\bar{Y}}(x|\bar{Y} = \bar{y}, X > 0, \bar{Y} > 0). \quad (9)$$

To construct the conditional predictive distribution $G_{X|\bar{Y}}$ according to its definition in Equation (8) requires five elements: $p_{01}, p_{11}, H_{\bar{Y}}(\bar{y}), F_{\bar{Y}}(\bar{y}), H_{X|\bar{Y}}$.

Introducing

$$\psi(\bar{y}) = \frac{p_{01} h_{\bar{Y}}(\bar{y}|X = 0, \bar{Y} = \bar{y} > 0)}{p_{11} f_{\bar{Y}}(\bar{y}|X > 0, \bar{Y} = \bar{y} > 0)} \quad (10)$$

Equation (8) can be rewritten as:

$$G_{X|\bar{Y}}(x|\bar{Y} = \bar{y}, \bar{Y} > 0) = \frac{\psi(\bar{y})}{1 + \psi(\bar{y})} + \frac{1}{1 + \psi(\bar{y})} H_{X|\bar{Y}}(x|\bar{y}) = \frac{\psi(\bar{y}) + H_{X|\bar{Y}}(x|\bar{y})}{1 + \psi(\bar{y})} \quad (11)$$

Our aim is to estimate $G_{X|\bar{Y}}(x|\bar{Y} = \bar{y}, \bar{Y} > 0)$ by modelling the conditional distribution $H_{X|\bar{Y}}$ through the UMCP processor that allows considering ensemble forecasts in the above-described bivariate dependence structure and takes advantage of relationships in the normal space.

3.2. Univariate model conditional processor (UMCP)

Based on the properties of the multi-normal distribution, [Todini \(2008\)](#) developed the MCP for producing a full predictive probability distribution of the variable of interest that conveniently handles multivariate situations, making it suitable for multi-model, multisite, and multi-lead time problems of streamflow predictions. MCP provides a predictive density defined as the probability density of a future occurrence given (conditional to) all the available knowledge, encapsulated in one or more model forecasts ([Krzysztofowicz 1999](#)). MCP, which under the assumption of the lag-1 Markov process made by [Krzysztofowicz \(1999\)](#) was proven ([Todini 2013](#)) to lead exactly to the same results of the HUP post-processor, allows for easier extensions to the multivariate problems arising from multi-model and multi-time forecasts.

The original MCP approach implies homoscedasticity and was further extended from the normal to the truncated normal distributions in order to account for heteroscedasticity ([Coccia & Todini 2011](#)). So far, the MCP has been applied to the derivation of predictive densities for continuous random variables such as river stage and discharge, monthly average precipitation and temperature ([Reggiani et al. 2016](#)), and recently extended to multivariate mixed binary continuous variables, primarily daily precipitation ([Massari et al. 2019](#); [Reggiani & Byoko 2019](#)). [Biondi & Todini \(2018\)](#) introduced a univariate version of the MCP designed for the analysis of ensembles of streamflow forecasts that does not require the reordering of the members and allows for an adequate assessment of the time heteroscedasticity of the estimated predictive distribution variance. Specifically, the predictive distribution is obtained by using the expected value of the ensemble as the predictor, and assuming the ensemble mean being affected, at each point in time, by a different estimation error, whose variance is derived from the ensemble spread. UMCP takes advantage of the flexibility of the meta-Gaussian approach that allows any form of marginal distribution functions and for a closed form of the resulting predictive distribution. The fundamentals of the UMCP are described in the following. For more details, the reader can refer to [Biondi & Todini \(2018\)](#).

Given the marginal conditional distribution $F(x)$, the observations x_t with $t \in 1, \dots, T$ where T is the length of the observations record $x_t > 0$, are converted into the Gaussian space via the NQT to obtain η_t the normal image of the observations. Similarly, the $m \times T$ ensemble predictions, $y_{i,t}$ with $t \in 1, \dots, T$ and $i \in 1, \dots, m$ where m is the number of ensemble members, are converted into the Gaussian space to obtain $\xi_{i,t}$, the normal image of the ensemble forecasts. All the ensemble members are converted together to underline that no a priori preference is set on individual ensemble members, which are assumed equiprobable and no reordering is required. In the Gaussian space, the converted variables will result into standard normal variables, namely, $\eta_t \sim N(0, 1)$ and $\xi_{i,t} \sim N(0, 1)$.

As previously mentioned, the UMCP considers the mean value of the ensemble forecasts as the predictor; in the normal space, the ensemble mean, $\bar{\xi}_t$, is defined as:

$$\bar{\xi}_t = \frac{1}{m} \sum_{i=1}^m \xi_{i,t} \quad (12)$$

with zero mean $\mu_{\bar{\xi}_t} = 0$ and variance $\sigma_{\bar{\xi}_t}^2 = \text{Var}\{\bar{\xi}_t\}$.

In order to account for the uncertainty on the ensemble mean, the UMCP considers $\bar{\xi}_t$ as a deterministic predictor affected by measurement errors represented by the ensemble spread at each step:

$$R_{t,\bar{\xi}_t} = \frac{1}{m-1} \sum_{i=1}^m (\xi_{i,t} - \bar{\xi}_t)^2 \quad (13)$$

Based on the properties of the multivariate normal distribution (Mardia *et al.* 1979), the MCP algorithm was modified into a Deming (1943) type regression in the standard normal space, providing the UMCP estimator of the conditional mean and variance of the predictive density of η_t given $\bar{\xi}_t$ as follows:

$$\begin{cases} \mu_{\eta_t|\bar{\xi}_t} = \frac{\gamma_{\eta\bar{\xi}_t}}{\sigma_{\bar{\xi}_t}^2 + R_{t,\bar{\xi}_t}} \bar{\xi}_t \\ \sigma_{\eta_t|\bar{\xi}_t}^2 = 1 - \frac{\gamma_{\eta\bar{\xi}_t}^2}{\sigma_{\bar{\xi}_t}^2 + R_{t,\bar{\xi}_t}} \end{cases} \quad (14)$$

where $\gamma_{\eta\bar{\xi}_t} = 1/(T-1) \sum_{t=1}^T \eta_t \bar{\xi}_t$ is the covariance of the observations and the ensemble mean and $R_{t,\bar{\xi}_t} = R_{t,\xi_t}/m$. The conditional variance, as $R_{t,\bar{\xi}_t}$ varies at each step, is not constant and residuals are heteroscedastic.

It is also worth noting that assuming a joint multivariate cumulative distribution function of transformed data in the meta-Gaussian approach, corresponding to the well-known multivariate normal distribution, can also be viewed as a Gaussian copula. The meta-Gaussian approach has several properties that are noteworthy, for example its analytical tractability that greatly facilitates computational burden particularly when dealing with multiple dependent variables, but, as noted for the Gaussian copula, there are complex situations in which it may result as inappropriate to model the dependence of transformed variables by a simple covariance matrix (Renard & Lang 2007; Bárdossy & Pegram 2009). It is, therefore, necessary to check the adequacy of the dependence structure implied by the Gaussian model. However, although the suitability of the Gaussian dependence model has been proved to be useful within the observations range in many applications, another subtle problem may arise from the tail dependence properties of the variables that may have a non-negligible influence on the computation of probabilities related to extreme values. The multivariate Gaussian model implies the asymptotic independence of marginal values and presents limitation in preserving the high-order joint moments, thus the actual joint probability of observing very extreme values, outside the range of observations, can be higher than that suggested by the Gaussian multivariate model (Renard & Lang 2007; Dimitriadis & Koutsoyiannis 2018; Koutsoyiannis 2021). Among the different approaches that can be used to improve results, Coccia & Todini (2011) within the MCP framework divided the entire normal domain into sub-domains where truncated normal distributions (TNDs) can be used.

The post-processing approach implies that an ensemble of QPFs is available at each time step. Unfortunately, the Civil Protection Department provides only two forecasting bounds, which are associated with specific quantiles of the ensembles. For this reason, in order to use the UMCP approach, a two-parameter Gamma distribution was fitted to the mentioned QPF forecasting bounds and used to generate a more numerous ensemble with equiprobable members at each time step. We use the predictive distribution from the UMCP approach as an estimate of $H_{X|\bar{Y}}$ in the binary-continuous process of Equation (11) to derive the conditional CDF $G_{X|\bar{Y}}$.

3.3. Performance metrics

3.3.1. Assessment of forecast accuracy, bias and reliability

The accuracy of both raw QPFs and post-processed outcomes, considered as a measure of agreement between individual pairs of forecasts and observations, was carried out considering a traditional deterministic performance measure such as the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - x_t)^2} \quad (15)$$

where x_t is the observation at time t and y_t corresponds to a deterministic predictor.

The bias, which measures the discrepancy between the mean of forecasts and the mean of observations is estimated as:

$$BIAS = \frac{1}{n} \sum_{t=1}^n (y_t - x_t) \quad (16)$$

Considering the raw QPFs, provided at each time step in the form of lower and upper bounds, their mean value or one of the bounds can be assumed as deterministic predictor.

When dealing with post-processors providing a probabilistic forecast of continuous predictands, a single deterministic value, in this case the expected value of the predictive distribution, is considered as predictor in Equations (15) and (16).

Nevertheless, to allow a complete and fair evaluation of a probabilistic forecast performance, the use of other appropriate diagnostic approaches is recommended. To assess reliability of probabilistic forecasts, with the aim of verifying the statistical consistency between the observed time series of the predictand and its predictive distribution (Gneiting *et al.* 2007), Laio & Tamea (2007) proposed the use of a diagnosis plot based on the probability integral transform (PIT). The approach plots transformed values $z_t = \Phi_t(x_t)$, i.e., the predictive CDF Φ calculated for observed values x_t , against their empirical CDF; if a forecasting system is reliable, the uniform distribution $U(0, 1)$ is appropriate for the constructed sample of z and the plot follows the 1:1 line. Moreover, analysing the shape of the resulting curve provides interesting insights about the properties of the predictive distribution (i.e., problems related to under or over dispersion, positive or negative bias).

In addition, the overall quality of the probabilistic forecast of a continuous scalar predictand is assessed with the continuous ranked probability score (CRPS) (Matheson & Winkler 1976; Hersbach 2000) that encompasses information about reliability, resolution and uncertainty within a single scalar measure. The CRPS is defined as the quadratic measure of discrepancy between the predictive CDF, Φ , and the empirical CDF of the observation x_t , and is defined for a single forecasting occasion as:

$$CRPS(\Phi, x_t) = \int_0^{\infty} [\Phi(\varpi) - \mathbb{1}(\varpi \geq x_t)]^2 d\varpi \quad (17)$$

where $\mathbb{1}(\varpi \geq x)$ is the heaviside function with value 1 when $\varpi \geq x$ and 0 otherwise. *CRPS* has the dimension of the variable of interest and has a minimum value of zero, which is obtained only in the case of a perfect deterministic forecast. The average value \overline{CRPS} over time steps is then evaluated.

3.3.2. Assessment of forecast skills of probability thresholds

Finally, the added operational value of the proposed approach is assessed through the ability to forecast rainfalls that exceed the critical thresholds set for the WZV. In case of deterministic forecasts, for example, considering the average value of the QPFs, skills are evaluated from the contingency table (Mason 1982) of the four possible outcomes for a binary classifier: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Using values from the contingency table, it is possible to calculate the true positive rate (TPR), i.e., an estimate of the probability that a critical rainfall will be correctly forewarned ($TPR = TP/(TP + FN)$), and the false positive rate (FPR), i.e., how often an actual negative instance is classified as positive ($FPR = FP/(FP + TN)$). In the case of probability forecasts, the prediction of a severe rainfall occurrence is

translated into the definition of a threshold on the exceedance probability of the critical rainfall value (e.g., De Luca & Versace 2017). The optimal choice for the probability threshold is task specific and can be determined through an assessment of the economic value for decision-making (Mylne 2002) and, for example, taking the decision which corresponds to the minimum (or the maximum) expected value of a Bayesian cost utility function (Martina & Todini 2009). In this study, the optimal probabilistic threshold is determined assessing the ROC curves, which relate TPR and FPR evaluated for different exceedance probability thresholds. See Wilks (2005) for details on ROC curves.

Specifically, the threshold is determined as the probability value that minimizes the distance to perfect classification (D2PC) or radial distance:

$$D2PC = \sqrt{(1 - TPR)^2 + FPR^2} \quad (18)$$

i.e., as the point closest to the top left corner of the ROC space that optimizes TPR and FPR.

4. RESULTS

4.1. Comparing QPFs and observed rainfall

Figure 2 shows, for different durations, the scatter plots of the maximum values observed among the rain gauges of the WVZ on the day to which the forecast refers and the QPF ranges along with their mean values. In the following, \overline{QPF} refers to the QPF mean value and QPF range limits are noted as QPF_{up} and QPF_{low} , respectively. The available observations and \overline{QPF} generally show a very low level of correlation (Pearson correlation coefficients r ranging from 0.488 to 0.585), resulting in QPFs overestimating the observed low precipitation values and, conversely, highly underestimating the higher ones.

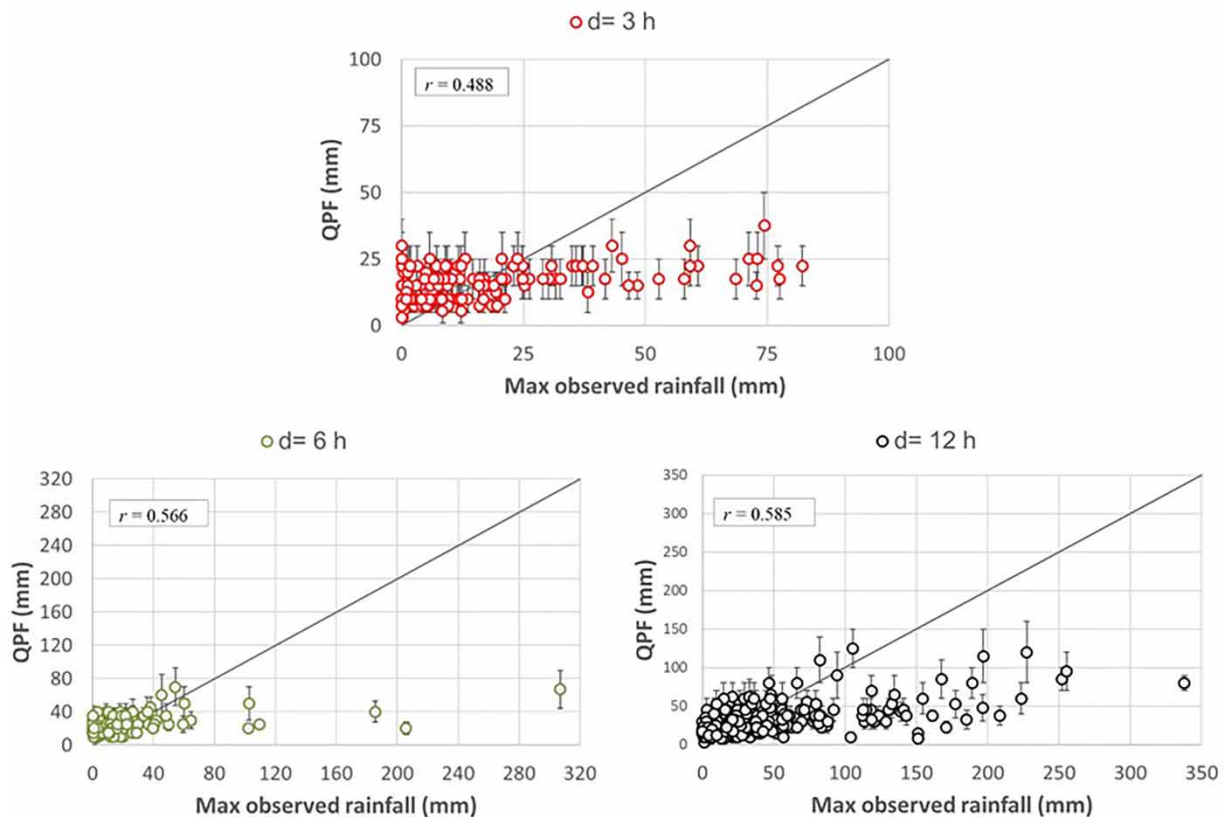


Figure 2 | Scatter plots of QPFs and maximum precipitation values observed on the day of the forecast among the rain gauges of the WVZ. Dots refer to the mean value of QPFs and whiskers indicate their range. r is the Pearson correlation coefficient between observations and \overline{QPF} .

4.2. Marginal distributions

As is apparent from Equations (6) and (7), marginal distribution functions of observations X and the predictor \bar{Y} can be constructed from the conditional distributions $F_X(x)$, $F_{\bar{Y}}(\bar{y})$ and $H_{\bar{Y}}(\bar{y})$, and the PoP occurrence for observed precipitation given a non-null prediction, p_{11} .

The continuous component of the mixed process, referring to the amount of precipitation, is here assumed Gamma distributed with shape parameter α and scale parameter β . Gamma distribution has been widely used for the probability distribution of precipitation amounts (e.g., Coe & Stern 1982; Wilks 1990; Hamill & Colucci 1998; Sloughter *et al.* 2007) as it can fit skewed data and is quite flexible (Heo *et al.* 2019).

Gamma distributions are fitted to datasets of observations and forecasts available for each duration. Moreover, the PoP occurrence, p_{11} , is estimated for each duration from the observations' sample corresponding to available QPF issuances.

Specifically, regarding forecasts, at each time step t , a Gamma distribution $F_t(y_t)$ is fitted to available QPF limits and their relative provided values of probability of exceedance (after November 2013 a probability of exceedance of 0.7 and 0.3 is assumed for QPF_{low} and QPF_{up} , respectively; the assumed probability values are the most observed in the available sample), and the resulting mean value and variance, consequently estimated. The dataset of QPF average values resulting from the Gamma fitting, that will be denoted as \overline{QPF}_{Ga} in the following, thus correspond to the \bar{y} that is further modelled with a Gamma distribution to estimate both the conditional CDFs $F_{\bar{Y}}(\bar{y})$ and $H_{\bar{Y}}(\bar{y})$. It is worth noting that the distribution parameters of $F_{\bar{Y}}(\bar{y})$ and $H_{\bar{Y}}(\bar{y})$ are estimated from the mean and the variance of \overline{QPF}_{Ga} (see Tables 2 and 3) that are, in turn, evaluated from the mean of the first and second order moments of the Gamma distributions $F_t(y_t)$ at each time step.

To accomplish the UMCP and to allow for a better representation of the ensemble spread, an ensemble of 100 members was generated at each time step according to the Gamma distributions $F_t(y_t)$ fitted to the quantile values provided by the Civil Protection Department; ensemble members were then grouped and analysed together in a single sample, QPF_{ens} , to avoid the reordering of the ensembles, under the assumption of exchangeability (Fraley *et al.* 2010).

Table 2 | Marginal conditional distributions of observed precipitations and QPFs, and PoP occurrence of observed precipitation

| Precipitation | Duration (h) | Sample | | Gamma distribution parameters | | PoP p_{11} |
|-----------------------|--------------|-----------|-----------|-------------------------------|--------------|--------------|
| | | Mean (mm) | S.D. (mm) | α (-) | β (mm) | |
| <i>Observed</i> | 3 | 18.42 | 20.44 | 0.772 | 23.863 | 0.884 |
| | 6 | 26.23 | 43.45 | 0.663 | 39.567 | 0.934 |
| | 12 | 48.66 | 54.78 | 0.828 | 58,747 | 0.980 |
| \overline{QPF}_{Ga} | 3 | 20.86 | 26.21 | 0.633 | 32.930 | |
| | 6 | 22.97 | 31.12 | 0.544 | 41.162 | |
| | 12 | 40.73 | 42.55 | 0.916 | 44.451 | |
| QPF_{ens} | 3 | 20.40 | 23.96 | 0.954 | 21.390 | |
| | 6 | 22.52 | 28.86 | 0.998 | 22.569 | |
| | 12 | 40.04 | 44.13 | 1.003 | 39.939 | |

Table 3 | Conditional distributions $H_{\bar{Y}}(\bar{y})$ of \overline{QPF}_{Ga} and conditional distributions for QPF_{ens}

| Precipitation | Duration (h) | Sample | | Gamma distribution parameters | |
|-----------------------|--------------|-----------|-----------|-------------------------------|--------------|
| | | Mean (mm) | S.D. (mm) | α (-) | β (mm) |
| \overline{QPF}_{Ga} | 3 | 17.82 | 28.93 | 0.379 | 46.967 |
| | 6 | 18.90 | 25.26 | 0.559 | 33.760 |
| | 12 | 21.00 | 12.99 | 2.613 | 8.035 |
| QPF_{ens} | 3 | 18.07 | 26.18 | 0.476 | 37.930 |
| | 6 | 18.60 | 16.91 | 1.279 | 14.532 |
| | 12 | 20.83 | 13.96 | 2.696 | 7.728 |

Table 2 summarizes sample statistics (i.e., mean and standard deviation (S.D.)) and distribution parameters of $F_X(x)$, $F_Y(\bar{y})$ estimated from samples of QPFs and strictly positive corresponding precipitation observations, as well as statistics and distribution parameters derived for the conditional distribution of the generated QPF_{ens} . Please note that the S.D. relevant to \overline{QPF}_{Ga} and QPF_{ens} is estimated as the square root of the total variance (Mahler & Dean 2001; Weiss 2006), given that when a forecast is issued at each time step in the form of a probability distribution, one cannot limit the estimation of the variance to the variance of the mean forecasts, but must estimate the total variance, namely, the sum of the variance of the mean forecast at each time step and the mean variance of all the time variant predictive densities.

Overall statistics of observations and precipitation forecasts suggest that \overline{QPF}_{Ga} underestimates the actual rainfall in the mean and generally shows lower variances with the exception of $d = 3$ h. Consequently, both the shape and the scale parameter values for observations differ from those of the forecasts. This statistical distinction, for the investigated example, can be traced back to the inadequacy of the forecasting process, to scaling issues and/or to the limited information gathered from the joint samples. Statistics of generated QPF_{ens} are, as expected, closer to those of \overline{QPF}_{Ga} .

Moreover, given the peculiarity of the case study, samples provide high values of PoP occurrence for all the investigated durations.

Similarly, the distribution $H_Y(\bar{y}) = H_Y(\bar{y}|X = 0, \bar{Y} > 0)$ for \overline{QPF}_{Ga} conditional on null observations, as well as the corresponding conditional distribution for QPF_{ens} are estimated; parameters and relative statistics are summarized in Table 3.

4.3. Accuracy and reliability performance of raw and post-processed forecasts

In order to assess forecast accuracy, in Table 4 we compare the RMSE of raw forecasts and of the post-processed outcomes. Regarding QPFs, RMSE is evaluated not only for \overline{QPF}_{Ga} , but also for both QPF_{up} and \overline{QPF} ; RMSE of post-processed forecasts is estimated considering the expected values of the predictive distribution $G_{X|Y}$. RMSE for the mean values of UMCP outcomes ($H_{X|\bar{Y}}$) back-transformed in the real space is also evaluated.

Results show that \overline{QPF} provides the best performance for $d = 3$ h, while RMSE values are found to be the highest for longer durations (i.e., $d = 6, 12$ h). Regarding unprocessed forecasts, QPF_{up} performs better for $d = 6, 12$ h. Accuracy metrics of QPF_{low} , not shown here, are significantly worse. \overline{QPF}_{Ga} shows performance similar to QPF_{up} for $d = 6$ and 12 h. RMSEs estimated in the real variable space reveal that the mean value of the predictive distribution $G_{\bar{X}|\bar{Y}}$ outperforms the best of the unprocessed forecasts for longer durations ($d = 6, 12$ h) and achieves performance comparable to \overline{QPF} for $d = 3$ h. RMSEs for the mean values of the UMCP-derived $H_{X|\bar{Y}}$, are found to be within the range of values obtained for all the other investigated deterministic predictors. It should be noted, that UMCP is used to estimate the conditional distribution defined and calibrated for the subset $X > 0, \bar{Y} > 0$.

The bias of \overline{QPF} , \overline{QPF}_{Ga} and of the expected value of the predictive distribution $G_{X|\bar{Y}}$ is shown in Table 5. It is evident that raw forecasts are more biased than post-processed forecasts. The bias for \overline{QPF} tends to be negative (e.g., -15.23 mm for $d = 12$ h), indicating a propensity to underestimate rainfall amounts. Bias of \overline{QPF}_{Ga} shows positive value for $d = 3$ h and confirms negative values, although less marked than \overline{QPF} , for longer durations. Post-processing almost reduces biases particularly for $d = 6$ and 12 h (e.g., -0.05 mm for $d = 12$ h).

By means of graphical representations in Figure 3, further insight is gained into the calibration and reliability of the probabilistic forecasts. For each forecasting occasion, PITs of the predictive distribution function $G_{X|\bar{Y}}$ calculated for the observed values, are plotted against their empirical cumulative distribution functions. Here, only strictly positive observations, which are the vast majority in the investigated samples, are considered. Unprocessed QPFs are also included in this analysis by assuming, at each step in time, as described above a Gamma distribution estimated from the QPF limits and the probability assigned to them. An ideal processor producing perfectly reliable forecasts should lead to a graph with markers lying on the

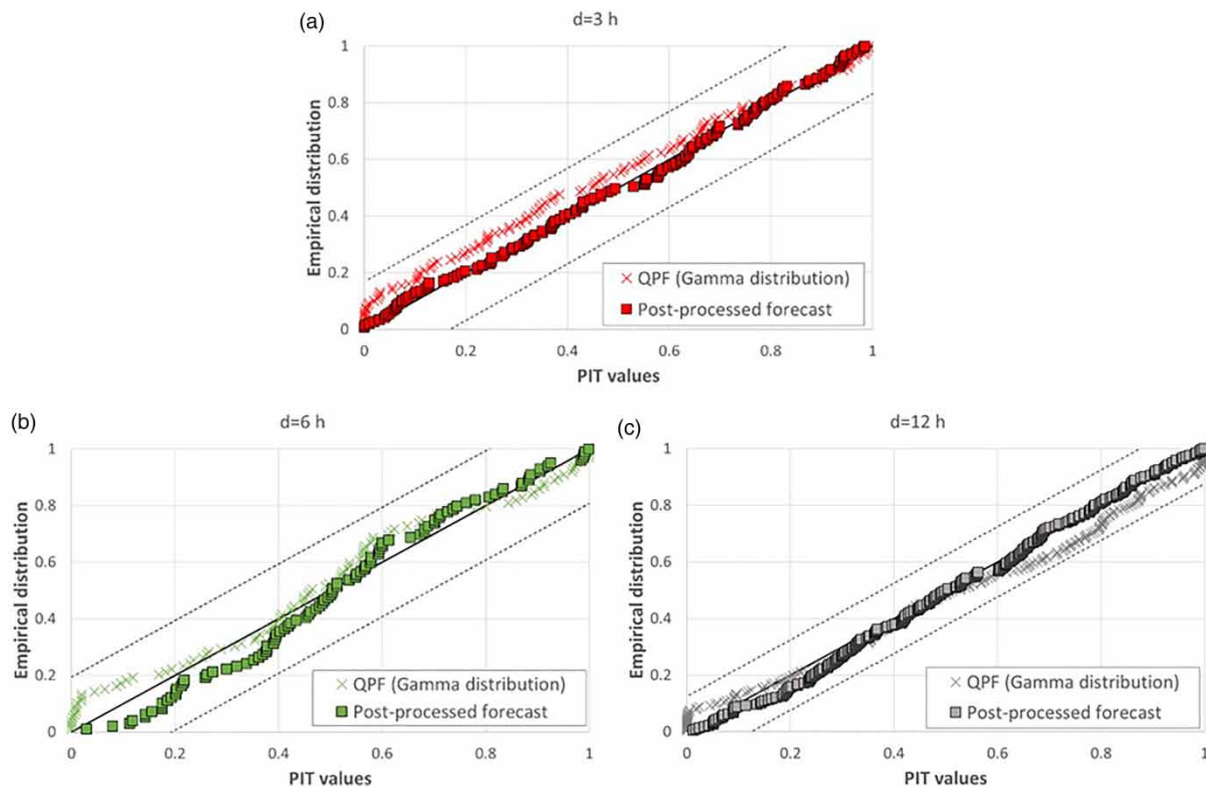
Table 4 | RMSE of QPFs and post-processed forecasts

| | <i>d</i> (h) | \overline{QPF} | QPF_{up} | \overline{QPF}_{Ga} | UMCP $H_{X \bar{Y}}$ | $G_{X \bar{Y}}$ |
|-----------|--------------|------------------|------------|-----------------------|----------------------|-----------------|
| RMSE (mm) | 3 | 17.82 | 18.21 | 19.74 | 18.25 | 17.94 |
| | 6 | 38.25 | 36.71 | 36.72 | 36.34 | 35.63 |
| | 12 | 48.19 | 45.06 | 45.70 | 46.30 | 44.74 |

\overline{QPF} , mean value of the QPF limits; QPF_{up} , upper limit of the QPF range; \overline{QPF}_{Ga} , mean value of the fitted Gamma distribution; $H_{X|\bar{Y}}$ UMCP, univariate model conditional processor; $G_{X|\bar{Y}}$, conditional predictive distribution.

Table 5 | BIAS and \overline{CRPS} of QPFs and post-processed forecasts

| | d (h) | \overline{QPF} | \overline{QPF}_{Ga} | $G_{X Y}$ |
|-------------------|---------|------------------|-----------------------|-----------|
| BIAS (mm) | 3 | -0.92 | 4.22 | 1.56 |
| | 6 | -6.70 | -1.79 | 0.50 |
| | 12 | -15.23 | -7.34 | -0.05 |
| \overline{CRPS} | 3 | | 9.38 | 8.78 |
| | 6 | | 13.86 | 13.40 |
| | 12 | | 22.61 | 21.52 |

**Figure 3** | Probabilistic integral transform (PIT) plots of post-processed forecasts and Gamma distributed QPFs for (a) $d = 3$ h; (b) $d = 6$ h; (c) $d = 12$ h. The black solid line is the bisector; black dashed lines represent the 5% Kolmogorov confidence bands.

bisector (i.e., the black solid line corresponding to a theoretical uniform distribution). Gamma distributions assumed for the two-member ensembles deviate from the ideal behaviour although points fall within the Kolmogorov bands (black dashed lines). A modest positive bias is detected at $d = 3$ h (Figure 3(a)), as apparent from the fact that the points lie above the bisector of the diagram, while a slight under-dispersive behaviour is recognized for $d = 6$ and 12 h, i.e., provide narrow predictions around the central values. Conversely, predictive distributions obtained via post-processing are always within the 5% Kolmogorov significance bands but closer to the bisector. The diagrams thus show that the reliability of probability forecasts is improved by post-processing QPFs and both underdispersion and frequency bias are reduced.

The results can be better interpreted looking at the predictive distributions in Figure 4. Figure 4(a), 4(c) and 4(e) depict $d = 3$, 6 and 12 h, the median of the predictive distribution (red line) and the 50% as well as the 90% uncertainty bands (dark and light grey bands, respectively) in the real space, along with observed precipitation. For the same rainfall durations, Figure 4(b), 4(d) and 4(f) compare observed precipitation with QPF ranges (full coloured bands) and with the 90% credible intervals and the expected value of the Gamma distribution fitted to the forecasts.

First, we observe that predictive distribution in the real space assumes the shape of a skewed Gamma-type distribution, as well as the QPF credible intervals derived from the Gamma assumption for the distribution of forecasts at each occasion; it is

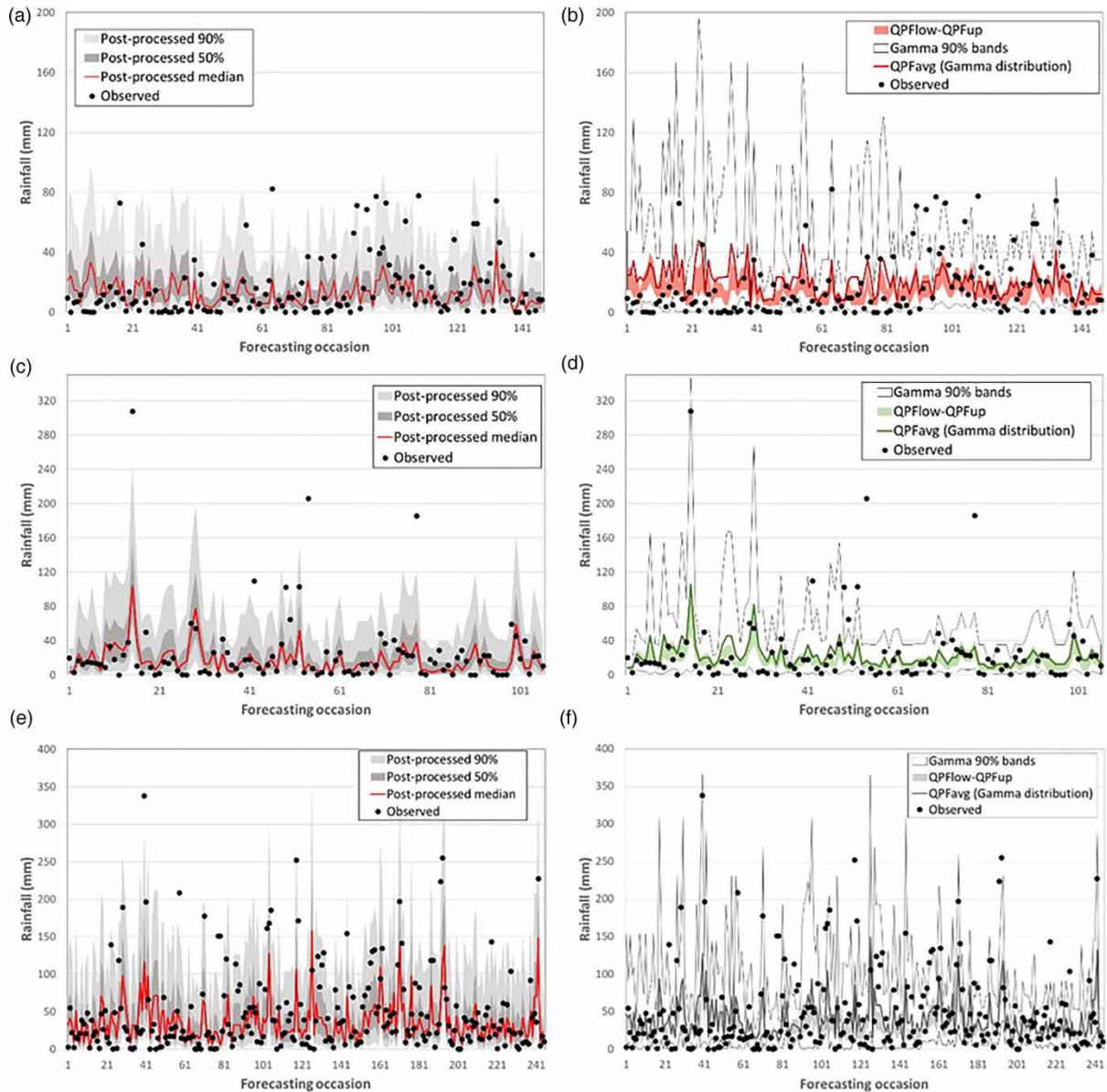


Figure 4 | Observed precipitation, median and credible intervals (50 and 90%) for the post-processed predictive distribution for (a) $d = 3$ h; (c) $d = 6$ h; (e) $d = 12$ h. Observed precipitation, QPF ranges, \overline{QPF}_{Ga} values and 90% credible intervals of the Gamma distribution for (b) $d = 3$ h; (d) $d = 6$ h; (f) $d = 12$ h. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/nh.2021.045>.

worth noting that, by construction, \overline{QPF}_{Ga} may fall outside the QPF limits. The width of the uncertainty bands of the QPF Gamma distribution reveals a higher sharpness for $d = 6$ and 12 h compared to $d = 3$ h confirming the under-dispersive behaviour shown in PIT plots of Figure 3(b) and 3(c). If, at least in principle, narrow confidence bands can be considered as a good property, the sharpness of the distribution does not give information about the correspondence between forecast and observations and are mitigated in this case by less satisfactory reliability and observed bias. The percentage of observations falling within the interquartile range (IQR) of QPF Gamma distribution is 39% for $d = 3$ h, 49% for $d = 6$ h and 44% for $d = 12$ h; percentages for observation included within the IQR of post-processed predictive distribution significantly increase to 45% for $d = 3$ h, 56% for $d = 6$ h and 52% for $d = 12$ h.

Mean CRPS values in Table 5 confirm that post-processed probabilistic forecasts yield better results than Gamma distribution fitted to QPFs limits. \overline{CRPS} of QPFs increases with duration, varying from 9.38 for $d = 3$ h to 22.61 for $d = 12$ h. The proposed post-processing approach leads to a slight improvement in \overline{CRPS} , that shows the same tendency to increase

with duration from 8.78 for $d = 3$ h to 21.52 for $d = 12$ h. The bias and the sharpness of the distribution, as for QPFs, generally increase the difference between associated probability distributions and heaviside functions and consequently result in higher CRPS values.

4.4. Skills in forecast of critical rainfall occurrence

The ability of the PQPF derived with the proposed post-processor of being used to identify the occurrence of potentially critical rainfall in the day of issuance is also tested. Moving to a probability perspective, an event is defined when the predictive probability of exceedance of an assumed critical rainfall exceeds a defined probability level. We choose to consider as critical, and indicate it with x_{cr} , the rainfall value associated with the lowest critical rainfall threshold (ordinary level) over the alert zone (Table 6). It should be noted that the events exceeding higher rainfall thresholds are a subset of the events exceeding this threshold.

The first step consisted in establishing the probability threshold value, suited to identify critical rainfall, through a ROC analysis. In the case of probabilistic forecasts, to be able to calculate associated TPR and FPR, a probability threshold is considered as the classifier to discriminate between event and non-event for the forecasts. To evaluate the requested performance statistics, forecast binary outcomes have been compared with the actual occurrences of the overcoming of x_{cr} by the observed rainfall.

The ROC curves in Figure 5(a) for $d = 3, 6, 12$ h plot TPR versus FPR metrics for a set of possible thresholds, obtained by assuming different $p = P(X \geq x_{cr})$, i.e., different values for the probability of exceedance of the defined critical rainfall (in this application p varies from 0.05 to 0.6). Here, we consider D2PC defined at Equation (18) to achieve the trade-off between correct detection and false alarms. Solid circles on the ROC curves highlight points corresponding to the minimum D2PC values obtained among the investigated probability thresholds. Optimal D2PC values at each duration are found for different values of p that hence define the probabilistic threshold p_{cr} : specifically, $p_{cr} = 0.10$ for $d = 3$ h, $p_{cr} = 0.15$ for $d = 6$ h and $p_{cr} = 0.25$ for $d = 12$ h.

We also evaluated the skill performances for the case of a deterministic forecast considering \overline{QPF}_{Ga} as predictor, and calculated TPR as well as FPR directly from the contingency table obtained by comparing \overline{QPF}_{Ga} and observed rainfall with the rainfall threshold x_{cr} . The resulting statistics are shown in Table 6 and evidenced as square symbols in the ROC space of Figure 5(b). Both TPR and FPR show low values, with TPR only slightly exceeding FPR.

As evident from performance metrics in Table 6, assuming a probability threshold as binary classifier significantly improves skill statistics of forecasting critical rainfalls. Using probabilistic forecasts, TPR values are strongly improved leading to a

Table 6 | Performance of post-processed probabilistic forecasts evaluated considering a probability threshold p_{cr} as binary classifier and performance for deterministic Gamma fitted and raw forecasts compared to the rainfall threshold x_{cr}

| | d (h) | p_{cr} | TP | FN | FP | TPR | FPR | D2PC |
|------------------------|---------|---------------|----|----|----|------|------|------|
| Probabilistic forecast | 3 | 0.1 | 14 | 3 | 34 | 0.82 | 0.26 | 0.32 |
| | 6 | 0.15 | 6 | 4 | 27 | 0.60 | 0.28 | 0.49 |
| | 12 | 0.25 | 39 | 17 | 48 | 0.70 | 0.25 | 0.39 |
| | | x_{cr} (mm) | TP | FN | FP | TPR | FPR | DPC |
| \overline{QPF}_{Ga} | 3 | 42 | 1 | 16 | 5 | 0.06 | 0.04 | 0.94 |
| | 6 | 53 | 2 | 8 | 0 | 0.20 | 0.00 | 0.80 |
| | 12 | 66 | 17 | 39 | 16 | 0.30 | 0.08 | 0.70 |
| | | x_{cr} (mm) | TP | FN | FP | TPR | FPR | DPC |
| \overline{QPF} | 3 | 42 | 0 | 17 | 0 | 0.00 | 0.00 | 1.00 |
| | 6 | 53 | 1 | 9 | 0 | 0.10 | 0.00 | 0.90 |
| | 12 | 66 | 12 | 44 | 1 | 0.21 | 0.01 | 0.79 |
| | | x_{cr} (mm) | TP | FN | FP | TPR | FPR | DPC |
| QPF_{up} | 3 | 42 | 1 | 16 | 0 | 0.06 | 0.00 | 0.94 |
| | 6 | 53 | 2 | 8 | 1 | 0.20 | 0.01 | 0.80 |
| | 12 | 66 | 19 | 37 | 16 | 0.34 | 0.18 | 0.68 |

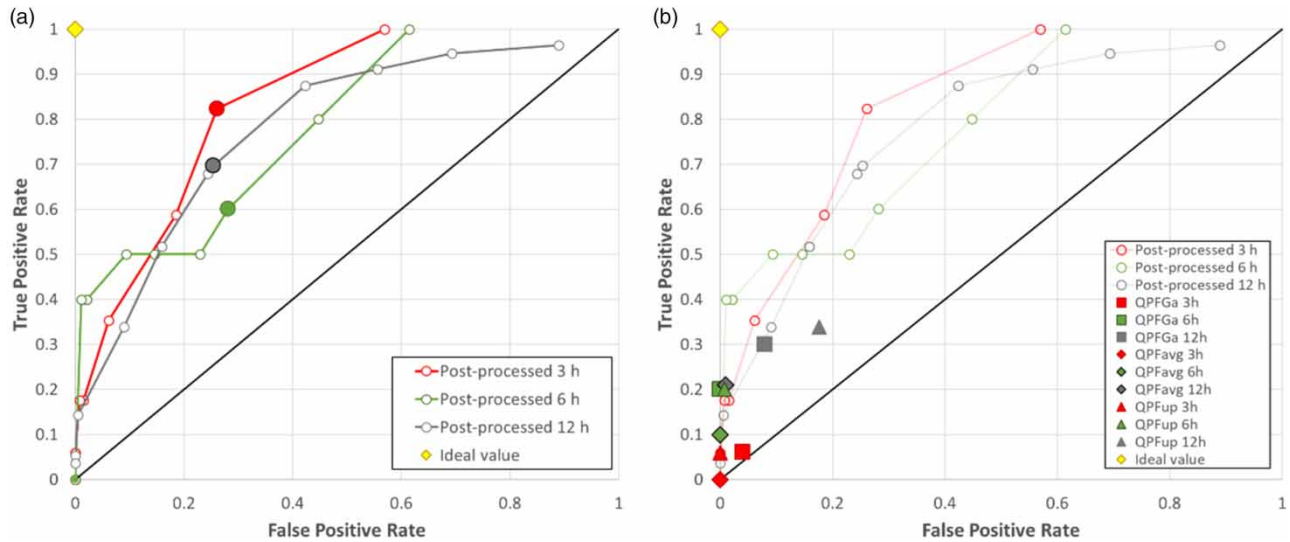


Figure 5 | (a) ROC curves for different durations obtained varying the probability of exceedance threshold of the critical rainfall. Circles highlight the best performing probability thresholds in terms of D2PC. (b) Performance of deterministic forecasts in the ROC plane (FPR, TPR). Square symbols correspond to TPR and FPR values obtained comparing QPF_{Ga} with rainfall thresholds. Diamonds and triangles refer to QPF and QPF_{up} of the raw deterministic forecasts, respectively.

moderate increase in corresponding FPR (always ≤ 0.3). Exemplarily, when considering the \overline{QPF}_{Ga} as deterministic forecasts for a duration of 6 hours, TPR is 0.20 (of all observed rainfall threshold exceedances, only 20% are forecasted), while FPR is 0 (no event is forecasted when no event is observed). By using probabilistic forecasts, a much higher TPR, equal to 0.6, can be achieved at the same time keeping FPR at 0.28. This means that 60% of all events are correctly forecasted, which poses a great improvement, particularly dealing with early warning systems. Performances are evaluated also for QPF and QPF_{up} of the raw forecasts; compared to \overline{QPF}_{Ga} only a slight improvement in terms of TPR and D2PC is found for QPF_{up} at $d = 12$ h as is shown in Table 6 and by triangles in Figure 5(b).

5. SUMMARY AND CONCLUSIONS

The presented case study shows a framework aimed at testing a post-processing approach for ensemble precipitation forecasts and uncertainty assessment, in the operational context of a WVZ of the national warning system in southern Italy. Specifically, we propose a parsimonious and computationally efficient methodology based on the inclusion of the UMCP in the mixed binary-continuous precipitation process for the derivation of the predictive uncertainty associated with multiple QPFs. It is worth noting that the analyses carried out on QPF statistics do not refer to the effectiveness of the alert system, whose warnings rely also on several pieces of other information, and only represent a case study to test the performance of the approach as well as potential implications for operational forecasting of critical rainfall and rainfall-induced damaging events.

Our main purpose was to assess the ability and the robustness of the proposed post-processing technique to be accurate and to reliably quantify forecast uncertainty. Evaluating a reliable predictive uncertainty distribution is challenging (Berthet *et al.* 2020) and is crucial in decision-making scenarios in order to set up robust Bayesian decision-making approaches (see, for example, Martina *et al.* (2006)). Overall, the results suggest that post-processing not only shows advantages in terms of accuracy (measured by RMSE) when considering the expected value of the predictive distribution, but can yield satisfactory correction of biases in raw forecasts and generate reliable QPFs suited to further generate coherent ensemble members. In addition, looking at the overall quality, the probabilistic forecasts proved to provide enhanced average CRPS values. The proposed approach can be further extended by using the post-processor in a multi-temporal framework allowing to account for the time dependence of the forecasting errors and to derive a representation of the predictive density as the joint probability of precipitation at different future times conditional on the forecasts for the same time intervals. The multi-temporal version of the MCP approach and results of its applications to flow forecast are presented in Barbetta *et al.* (2017). Here, the lack of sufficiently long time series of QPFs issued for different time intervals prevented the application

of the multi-temporal approach and future work is envisaged to assess its skill in the post-processing of rainfall forecasts. It is likewise important to stress the implications of some assumptions (e.g., the use of light tail distributions) of the methodology that need to be verified on longer time series of observations and that can lead to an underestimation of the probability of extreme events.

Moreover, the use of a predictive distribution that accounts for the inherent uncertain nature of the forecast (Montanari & Koutsoyiannis 2012) requires a change of perspective in the warning threshold definition to thoroughly benefit from probabilistic forecasts. To take full advantage of the probabilistic forecasts, one should revert from the commonly used deterministic to probabilistic thresholds, as advocated by Todini (2017, 2018). In this perspective, the predictive ability of the post-processed PQPF was compared to that of raw forecasts by switching from the classic ‘deterministic threshold paradigm’ to the ‘probabilistic threshold paradigm’. The ‘deterministic’ rainfall threshold, represented by a rainfall critical value, is de facto substituted by a threshold on the probability of exceedance of a critical rainfall value. For the tested instances, assuming a probability threshold as binary classifier to determine the occurrence of potentially critical rainfall, improved skill statistics compared to the standard use of deterministic rainfall forecasts and deterministic thresholds. Particularly, the tested approach improved the ability to correctly forecast overcoming the critical rainfall, i.e., enhanced TPR values at the cost of a moderate increase in corresponding FPR. Overall, results confirm the benefit of post-processing forecasts in early warning systems, reinforced the perception that moving to the ‘probabilistic threshold paradigm’ can increase the accuracy of the alerts and may inspire its use as a valuable approach in predicting imminent or developing emergencies and disasters.

ACKNOWLEDGEMENTS

This work was partially carried out within the cooperation agreement between Italian National Civil Protection and Camilab-DIMES Unical on ‘Lo sviluppo della conoscenza, delle metodologie e delle tecnologie utili alla realizzazione, presso i centri funzionali, di sistemi di monitoraggio, previsione e sorveglianza nazionali, nonché per l’attuazione dell’organizzazione della funzione di supporto tecnico – scientifico nell’ambito del Servizio Nazionale della protezione civile. sviluppo, a beneficio della rete dei centri funzionali, di prodotti operativi finalizzati alla stima e alla previsione delle precipitazioni sul territorio nazionale’.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

- Barbetta, S., Coccia, G., Moramarco, T., Brocca, L. & Todini, E. 2017 [The multi temporal/multi-model approach to predictive uncertainty assessment in real-time flood forecasting](https://doi.org/10.1016/j.jhydrol.2017.06.030). *Journal of Hydrology* **551**, 555–576. <https://doi.org/10.1016/j.jhydrol.2017.06.030>.
- Bárdossy, A. & Pegram, G. G. S. 2009 [Copula based multisite model for daily precipitation simulation](https://doi.org/10.1016/j.jhydrol.2009.06.014). *Hydrology and Earth System Sciences* **13**, 2299–2314.
- Bell, T. H. 1987 [A space-time stochastic model of rainfall for satellite remote-sensing studies](https://doi.org/10.1029/JD092iD08p09631). *Journal of Geophysical Research* **92**, 9631–9643. doi.org/10.1029/JD092iD08p09631.
- Berthet, L., Bourgin, F., Perrin, P., Viatgé, J., Marty, R. & Piotte, O. 2020 [A crash-testing framework for predictive uncertainty assessment when forecasting high flows in an extrapolation context](https://doi.org/10.1016/j.jhydrol.2020.07.014). *Hydrology and Earth System Sciences* **24**, 2017–2041.
- Biondi, D. & Todini, E. 2018 [Comparing hydrological postprocessors including ensemble predictions into full predictive probability distribution of streamflow](https://doi.org/10.1029/2017WR022432). *Water Resources Research* **54** (12). doi.org/10.1029/2017WR022432.
- Bremnes, J. B. 2004 [Probabilistic forecasts of precipitation in terms of quantiles using NWP model output](https://doi.org/10.1175/JCLI14030338). *Monthly Weather Review* **132**, 338–347.
- Cloke, H. L. & Pappenberger, F. 2009 [Ensemble flood forecasting: a review](https://doi.org/10.1016/j.jhydrol.2009.06.014). *Journal of Hydrology* **375**, 613–629.
- Coccia, G. & Todini, E. 2011 [Recent developments in predictive uncertainty assessment based on the model conditional processor approach](https://doi.org/10.1016/j.jhydrol.2011.06.014). *Hydrology and Earth System Sciences* **15**, 3253–3274. <https://doi.org/10.1016/j.jhydrol.2011.06.014>.
- Coe, R. & Stern, R. D. 1982 [Fitting models to daily rainfall data](https://doi.org/10.1175/1520-0476(1982)21%3E1024%3E3:0%3E::AID-JAPM1024%3E3.0.CO;2). *Journal of Applied Meteorology* **21** (7), 1024–1031.
- De Luca, D. L. & Versace, P. 2017 [Diversity of rainfall thresholds for early warning of hydro-geological disasters](https://doi.org/10.1016/j.jhydrol.2017.06.014). *Advances in Geosciences* **44**, 53–60.
- Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J. & Ramos, M. H. 2007 [Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting](https://doi.org/10.1016/j.jhydrol.2007.06.014). *Environmental Hazards* **7** (2), 115–127.
- Deming, W. E. 1943 *Statistical adjustment of data*. Wiley, NY (Dover Publications edition, 1985). ISBN 0-486-64685-8.

- Dimitriadis, P. & Koutsoyiannis, D. 2018 Stochastic synthesis approximating any process dependence and distribution. *Stochastic Environmental Research & Risk Assessment* **32** (6), 1493–1515. doi:10.1007/s00477-018-1540-2.
- Evin, G., Favre, A. C. & Hingray, B. 2018 Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences* **22**, 655–672.
- Fraleigh, C., Raftery, A. E. & Gneiting, T. 2010 Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Monthly Weather Review* **138**, 190–201. doi:10.1175/2009MWR3046.1.
- Glahn, H. R. & Lowry, D. A. 1972 The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology* **11**, 1203–1211.
- Gneiting, T., Raftery, A. E., Westveld, A. H. & Goldman, T. 2005 Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review* **133**, 1098–1118.
- Gneiting, T., Balabdaoui, F. & Raftery, A. E. 2007 Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc. Ser. B* **69**, 243–268.
- Hamill, T. M. & Colucci, S. J. 1998 Evaluation of Eta–RSM ensemble probabilistic precipitation forecasts. *Monthly Weather Review* **126**, 711–724.
- Hamill, T. M. & Whitaker, J. S. 2006 Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Monthly Weather Review* **134**, 3209–3229.
- Hamill, T. M., Whitaker, J. S. & Wei, X. 2004 Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review* **132**, 1434–1447.
- Hemri, S., Lisniak, D. & Klein, B. 2015 Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research* **51**, 7436–7451.
- Heo, J.-H., Ahn, H., Shin, J.-Y., Kjeldsen, T. R. & Jeong, C. 2019 Probability distributions for a quantile mapping technique for a bias correction of precipitation data: a case study to precipitation data under climate change. *Water* **11**, 1475. <https://doi.org/10.3390/w11071475>.
- Herr, H. D. & Krzysztofowicz, R. 2005 Generic probability distribution of rainfall in space: the bivariate model. *Journal of Hydrology* **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. *Weather 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).
- Kelly, K. S. & Krzysztofowicz, R. 2000 Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* **36**, 2643–2653. <https://doi.org/10.1029/2000WR900061>.
- Koutsoyiannis, D. 2006 An entropic-stochastic representation of rainfall intermittency: the origin of clustering and persistence. *Water Resources Research* **42** (1), W01401. doi:10.1029/2005WR004175.
- Koutsoyiannis, D. 2021 *Stochastics of Hydroclimatic Extremes – A Cool Look at Risk*. Kallipos, Athens, Greece.
- Koutsoyiannis, D., Yao, H. & Georgakakos, A. 2008 Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. *Hydrological Sciences Journal* **53** (1), 142–164.
- Krzysztofowicz, R. 1999 Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research* **35** (9), 2739–2750. <https://doi.org/10.1029/1999WR900099>.
- Krzysztofowicz, R. 2002 Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology* **268**, 16–40.
- Krzysztofowicz, R. 2008 Bayesian processor of ensemble: concept and development. In: *19th Conference on Probability and Statistics in the Atmosphere Sciences*, New Orleans, LA, USA.
- Krzysztofowicz, R. & Maranzano, C. J. 2006 *Bayesian Processor of Output for Probabilistic Quantitative Precipitation Forecasting*. American Geophysical Union, San Francisco, CA, USA.
- Laio, F. & Tamea, S. 2007 Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences* **11** (4), 1267–1277. <https://doi.org/10.5194/hess-11-1267-2007>.
- Li, W., Duan, Q., Miao, C., Ye, A., Gong, W. & Di, Z. 2017 A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdisciplinary Reviews Water* **4** (6). <https://doi.org/10.1002/wat2.1246>
- Madadgar, S., Moradkhani, H. & Garen, D. 2014 Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes* **28**, 104–122.
- Mahler, H. C. & Dean, C. G. 2001 Chapter 8: Credibility. In: *Foundations of Casualty Actuarial Science*, 4th edn. Casualty Actuarial Society, Arlington County, VA, USA, pp. 525–526. ISBN 978-0-96247-622-8.
- Mardia, K. V., Kent, J. T. & Bibby, J. M. 1979 *Multivariate Analysis. Probability and Mathematical Statistics*. Academic Press, London, UK.
- Martina, M. L. V. & Todini, E. 2009 Bayesian rainfall thresholds for flash flood guidance. In: *Flood Risk Management: Research and Practice* (Samuels, P., Huntington, S., Allsop, W. & Harrop, J., eds). Taylor & Francis Group, London, UK. ISBN 978-0-415-48507-4.
- Martina, M. L. V., Todini, E. & Libralon, A. 2006 A Bayesian decision approach to rainfall thresholds based flood warning. *Hydrology and Earth System Sciences* **10** (3), 413–426.
- Mason, I. 1982 A model for assessment of weather forecasts. *Aust. Meteor. Mag.* **30**, 291–303.
- Massari, C., Maggioni, V., Barbetta, S., Brocca, L., Ciabatta, L., Camici, S., Moramarco, T., Coccia, G. & Todini, E. 2019 Complementing near-real time satellite rainfall products with satellite soil moisture-derived rainfall through a Bayesian inversion approach. *Journal of Hydrology* **573**, 341–351.
- Matheson, J. E. & Winkler, R. L. 1976 Scoring rules for continuous probability distributions. *Management Science* **22**, 1087–1096. <https://doi.org/10.1287/mnsc.22.10.1087>.

- Mielke Jr., P. W. & Johnson, E. S. 1974 Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resources Research* **10** (2), 223–236.
- Montanari, A. & Koutsoyiannis, D. 2012 A blueprint for process-based modelling of uncertain hydrological systems. *Water Resources Research* **48**, W09555. doi:10.1029/2011WR011412.
- Mylne, K. R. 2002 Decision-making from probability forecasts based on forecast value. *Meteorological Applications* **9**, 307–315. doi:10.1017/S1350482702003043.
- Panofsky, H. W. & Brier, G. W. 1968 *Some Applications of Statistics to Meteorology*. The Pennsylvania State University Press, Philadelphia, PA, USA.
- Piani, C., Haerter, J. & Coppola, E. 2010 Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology* **99**, 187–192. doi:10.1007/s00704-009-0134-9.
- Raftery, A. E., Gneiting, T., Balabdaoui, F. & Polakowski, M. 2005 Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* **133**, 1155–1174.
- Reggiani, P. & Byoko, O. 2019 A Bayesian processor of uncertainty for precipitation forecasting using multiple predictors and censoring. *Monthly Weather Review* **147**, 4367–4387. doi: 10.1175/MWR-D-19-0066.1.
- Reggiani, P. & Weerts, A. H. 2009 A Bayesian approach to decision making under uncertainty: an application to real-time forecasting in the river Rhine. *Journal of Hydrology* **356**, 56–69.
- Reggiani, P., Coccia, G. & Mukhopadhyay, B. 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>.
- Reggiani, P., Todini, E., Boyko, O. & Buizza, R. 2021 Assessing uncertainty for decision-making in climate adaptation and risk mitigation. *International Journal of Climatology* **41**, 2891–2912.
- Renard, B. & Lang, M. 2007 Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Advances in Water Resources* **30**, 897–912.
- Robertson, D. E., Shrestha, D. L. & Wang, Q. J. 2013 Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrology and Earth System Sciences* **17**, 3587–3603. <https://doi.org/10.5194/hess-17-3587-2013>.
- Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X. & Seo, D. J. 2007 Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrology and Earth System Sciences Discussions* **4**, 655–717.
- Serinaldi, F. 2009 A multisite daily rainfall generator driven by bivariate copula-based mixed distributions. *Journal of Geophysical Research* **114**, D10103. doi.org/10.1029/2008JD011258.
- Slougher, J. M. L., Raftery, A. E., Gneiting, T. & Fraley, C. 2007 Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review* **135**, 3209–3220. <https://doi.org/10.1175/MWR3441.1>.
- Todini, E. 2008 A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management* **6**, 123–137.
- Todini, E. 2013 From HUP to MCP: analogies and extended performances. *Journal of Hydrology* **477**, 33–42.
- Todini, E. 2017 Flood forecasting and decision making in the new millennium. Where are we? *Water Resources Management* **31** (10), 3111–3129. doi:10.1007/s11269-017-1693-7.
- Todini, E. 2018 Paradigmatic changes required in water resources management to benefit from probabilistic forecasts. *Water Security* **3**, 9–17.
- Todini, E. & Di Bacco, M. 1997 A combined pölya process and mixture distribution approach to rainfall modeling. *Hydrology and Earth System Sciences* **1** (2), 367–378.
- Tsoukalas, I., Efstratiadis, A. & Makropoulos, C. 2019 Building a puzzle to solve a riddle: a multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. *Journal of Hydrology* **575**, 354–380. doi.org/10.1016/j.jhydrol.2019.05.017.
- United Nations Development Programme (UNDP) 2018 *Five Approaches to Build Functional Early Warning Systems*. Available from: <https://reliefweb.int/report/world/five-approaches-build-functional-early-warning-systems>
- Verkade, J. S., Brown, J. D., Reggiani, P. & Weerts, A. H. 2013 Postprocessing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology* **501**, 73–91.
- Wang, Q. J., Robertson, D. E. & Chiew, F. H. S. 2009 A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research* **45**, 641–648.
- Weerts, A. H., Winsemius, H. C. & Verkade, J. S. 2011 Estimation of predictive hydrological uncertainty using quantile regression: examples from the national flood forecasting system (England and Wales). *Hydrology and Earth System Sciences* **15** (1), 255–265.
- Weiss, N. A. 2006 *A Course in Probability*. Addison-Wesley, Boston, MA, USA.
- Wilks, D. S. 1990 Maximum likelihood estimation for the Gamma distribution using data containing zeros. *Journal of Climate* **3**, 1495–1501.
- Wilks, D. S. 2005 *Statistical Methods in the Atmospheric Sciences: An Introduction*, 2nd edn. International Geophysics Series, vol 59, Academic Press, Elsevier, Amsterdam.
- Wilks, D. S. 2009 Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications* **16**, 361–368.
- Wilks, D. S. & Hamill, T. M. 2007 Comparison of ensemble-MOS methods using GFS reforecasts. *Monthly Weather Review* **135**, 2379–2390. doi:10.1175/MWR3402.1.

First received 6 May 2021; accepted in revised form 20 September 2021. Available online 19 October 2021