Article

Predictive Uncertainty Estimation of Hydrological Multi-Model Ensembles Using Pair-Copula Construction

Bastian Klein [1,*], Dennis Meissner [1], Hans-Ulrich Kobialka [2] and Paolo Reggiani [3]

[1]   Department M2—Water Balance, Forecasting and Predictions, Federal Institute of Hydrology (BfG),
     Am Mainzer Tor 1, Koblenz 56068, Germany; meissner@bafg.de
[2]   Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Schloss Birlinghoven,
     Sankt Augustin 53757, Germany; hans-ulrich.kobialka@iais.fraunhofer.de
[3]   Department of Civil Engineering, University of Siegen, Paul-Bonatz-Str. 9-11, Siegen 57068, Germany;
     paolo.reggiani@uni-siegen.de
*   Correspondence: klein@bafg.de; Tel.: +49-261-1306-5256; Fax: +49-261-1306-5280

Abstract: Predictive uncertainty (PU) is defined as the probability of occurrence of an observed variable of interest, conditional on all available information. In this context, hydrological model predictions and forecasts are considered to be accessible but yet uncertain information. To estimate the PU of hydrological multi-model ensembles, we apply a method based on the use of copulas which enables modelling the dependency structures between variates independently of their marginal distributions. Given that the option to employ copula functions imposes certain limitations in the multivariate case, we model the multivariate distribution as a cascade of bivariate copulas by using the pair-copula construction. We apply a mixture of probability distributions to estimate the marginal densities and distributions of daily flow rates for various meteorological and hydrological situations. The proposed method is applied to a multi-model ensemble involving two hydrological and one statistical flow models at two gauge stations in the Moselle river basin. Verification and inter-comparison with other PU assessment methods show that copulas are well-suited for this scope and constitute a valid approach for predictive uncertainty estimation of hydrological multi-model predictions.

1. Introduction

Hydrological real-time forecasts are important means for decision support in water resources management and optimal control of rivers, for instance, for navigation or hydropower production. At the same time, forecasts are invaluable for flood risk management, as demonstrated during the major floods in Central Europe in June 2013. The benefit of an operational forecasting system depends foremost on the reliability and accuracy of the hydrological prediction. Excess non-quantified uncertainty reduces the benefits of predictions in the decision-making process.

Sources of uncertainty in the meteorological-hydrological forecasting chain are attributable to meteorological observations (*i.e.*, measurement and interpolation errors) and numerical weather forecasts (chaotic-deterministic behavior of atmospheric processes), model errors and parameter uncertainty as well as knowledge gaps on initial and boundary conditions of hydrological and hydraulic models used for flow prediction.

Rational decision-making on issues such as maximum vessel load, optimized operation of retention areas to shave flood peaks or timely release of water from reservoirs should optimally be based on objective cost-benefit analysis. This analysis depends on uncertain predictions of unknown quantities such as water level, flow rates or inflow volumes to reservoirs. The uncertainty of the unknown future value y can be characterized by a probability distribution conditioned on all information available at the onset of the forecast [1]. Ignoring the inherent uncertainty of the prediction, as in the case of purely deterministic forecasts, precludes truly rational decision-making.

In hydrological forecasting, the future flow rates or water levels are given by output of hydrological and hydraulic models. The Predictive Uncertainty (PU) of the observed variable at time t is the probability distribution $f(y_t | \hat{y}_t)$ of y conditional on a single or multiple model prediction(s) $\hat{y}$ [2]. The predictive uncertainty is conditional on the model structure, its parameters and the forcing data. The hydrological model predictions $\hat{y}$ can be a deterministic prediction by single or multiple models (multi-model ensemble) forced by a deterministic meteorological prediction or by meteorological ensemble predictions.

The definition of PU is often confused with the so called "validation uncertainty", which is the ability of a model to reproduce reality [3]. It describes the uncertainty $f(\hat{y}_t | y_t)$ of the prediction $\hat{y}$ conditional on the known observation y. This uncertainty is important to describe and improve the skill of hydrological models. What is more important to a flow forecaster is the uncertainty of the future observation given a particular decision than the uncertainty of the hydrological model prediction.

When dealing with uncertainty in operational practice the term has a negative connotation as an expression of "lack of knowledge", with the result that uncertain information is still ignored in many cases, while only the seemingly certain deterministic forecast is considered by operational users. To overcome this, well knowing that the future is not perfectly predictable, uncertainty should be rather communicated comprehensively as a measure of "predictive knowledge" [4]. To demonstrate this, a number of studies revealed the added potential economic value of using probabilistic forecasts in decision-making [4–6].

In recent years, a large number of different methods have been developed to estimate the PU of hydrological predictions. Most of the methods use meta-Gaussian models to describe dependency structures between random variables: prediction(s) *vs.* observation or prediction error *vs.* prediction, as it is much easier to build multivariate distributions when variables are Normal. Hydrologic variables such as flow rate and water level are generally non-Gaussian. Hence, variables are transformed to Normal space and the predicted Normal quantiles are then back-transformed to the real space by means of inverse transformation.

In earlier works on the topic of post-processing and PU assessment in hydrological forecasting, Krzysztofowicz [1,7] introduced the Bayesian Forecasting Framework BFS handling hydrological model uncertainty of a single model prediction by means of the Hydrological Uncertainty Processor HUP [8]. As prior density, a first-order Markov chain is used to model the river process, a representation which tends to oversimplify reality. A first-order Markov chain can be adequate to represent the recession curve but not necessarily the rising limb of the flood wave. Another drawback of the HUP is that linear regression models used for prior density and likelihood cannot easily handle error heteroscedasticity. Krzysztofowicz and Maranzano [9] extended the HUP to a multivariate version, which generalizes the univariate HUP to account for the temporal dependence between consecutive values. Reggiani *et al.* [10] applied the HUP, which was originally developed for deterministic forecasts, to estimate the PU in hydro-meteorological ensemble forecasts.

The Model Conditional Processor (MCP) [2] relies on a meta-Gaussian model to estimate the PU of the predictions by multiple hydrological models. To address heteroscedasticity in the error variance, Coccia and Todini [11] extended the MCP by using Multivariate Truncated Normal (MTN) distributions to model the joint multivariate probability distribution of multiple variates in the Normal space. In recent applications, the MCP is being extended to include hydro-meteorological ensemble forecasts [12].

Weerts *et al.* [13] applied Quantile Regression (QR) to consider heteroscedasticity in the error variance. QR estimates a separate regression relationship for each quantile. Depending on the number of forecast

models at least two parameters need to be estimated for each quantile, leading to a non-parsimonious problem of approximating the conditional probability distribution. A potential shortcoming arising with QR is that quantiles may cross due to their independent estimation. To mitigate this problem, Lopez *et al.* [14] applied a non-crossing QR variant. Depending on the distribution of residuals, QR not always works well with heteroscedastic error structures, a case frequently encountered in the peak flow range. In this context, Coccia and Todini [11] showed an example, where QR underperforms with respect to other methods.

Raftery *et al.* [15] proposed Bayesian Model Averaging (BMA) for PU estimation of multiple model predictions. Uncertainty is estimated as a weighted mean of the predictive distributions of individual models. In classical BMA, constant variance is assumed for the predictive distributions of individual models, irrespective of magnitude of the variable. The assumption of homoscedasticity of the error variance suffers from the shortcomings already mentioned for HUP with hydrological variables. BMA has been frequently applied in hydrology over recent years [16–18], while Madadgar & Moradkhani [18] introduced a combination of copulas and BMA.

Another method to post-process hydro-meteorological ensemble forecasts is Ensemble Model Output Statistics (EMOS) [19,20], a variant of Model Output Statistics (MOS) [21]. EMOS uses the variance and the weighted mean of ensemble predictions to estimate the uncertainty distribution. Madadgar *et al.* [22] applied copulas in post-processing of stream flow ensemble predictions.

All methods presented thus far are parametric. Brown and Seo [23] proposed a non-parametric post-processor which is similar to indicator co-kriging in geostatistics. Nevertheless, the accuracy of a non-parametric distribution remains highly dependent on the number of available observations and predictions.

Here, we propose to use the so-called pair-copula construction to build the multivariate probability distribution to estimate the PU of the predictions of three different prediction models: a conceptual, a physically-based and a statistical hydrologic response model. All three models are forced by observed meteorological data, implying that all hydrologic uncertainties described above are quantified by the PU assessment approach. The uncertainty of numerical weather prediction models has been eliminated, as observations are assumed to constitute "perfect" hindcasts.

The paper is structured as follows: first, we describe the study area and the models used. Then, we present the copula theory and the pair-copula construction as well as the mixed probability distributions used to estimate marginal distributions of the random variables. We also describe the methods applied as reference for uncertainty estimation and verification. Finally, we present the analysis and discuss the results. The paper ends with a conclusions section. All analyses have been performed using the statistical software R [24].

2. Materials and Methods

2.1. Study Area

The methods are applied to daily mean flow rates at two gauges in the Moselle river basin, which is an important tributary of the river Rhine. Two different catchment sizes are considered: gauge Bollendorf located at the river Sauer with an area of 3,222 km^2 and a mean flow of 40 m^3/s and gauge Trier at the Moselle with an area of 23,857 km^2 and a mean flow of 280 m^3/s (see Figure 1 and Table 1). Both gauges show a rainfall dominated (pluvial) flow regime with maximum flow during winter and early spring and minimum flow in late summer.

Table 1. Validation gauges, catchment area, mean flow, and maximum observed mean daily flow.

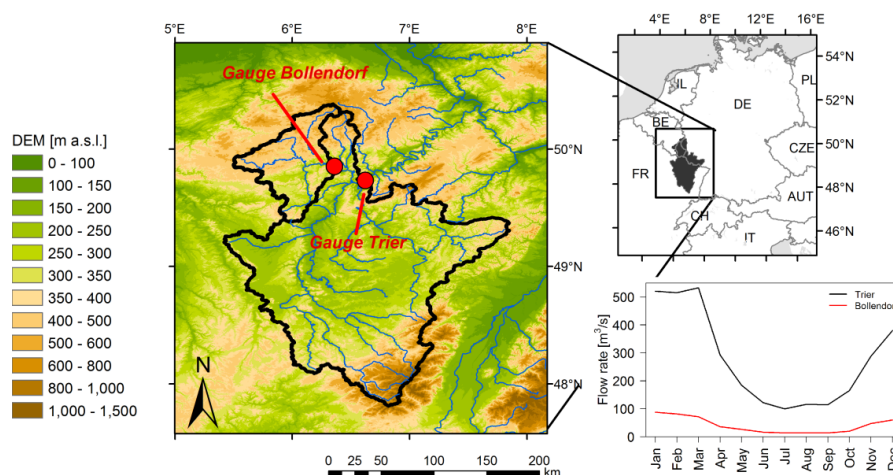| Gauge	River	Area (km^2)	Mean Flow (m^3/s)	Max. Flow (m^3/s)
Trier	Moselle	23,857	280	3840
Bollendorf	Sauer	3213	40	823

Figure 1. Study area Moselle river, elevation above sea level (digital elevation model DEM), sub-catchments gauge Trier and gauge Bollendorf, long-term mean monthly flow (1981–2010).

2.2. Hydrological and Statistical Models

Hourly runoff simulations from two hydrological and one statistical models aggregated to daily values are used to estimate the predictive uncertainty at the two gauges: the conceptual model HBV [25], the physically based Representative Elementary Watershed model REW [26,27] and the statistical model Echo State Networks (ESN) based on artificial Recurrent Neural Networks (RNN) [28]. All models use observed meteorological data, which are spatially interpolated to the subbasins of the specific models as meteorological forcing series.

For the semi-distributed HBV model, the catchment is subdivided into four (gauge Bollendorf) and respectively 26 (gauge Trier) subbasins. These subbasins are further subdivided into hydrological response units according to land use and elevation classes. The hydrological processes for runoff generation are calculated for those HRUs on the basis of linear conceptual stores. The model estimates runoff at hourly resolution using temperature and precipitation forcing time series that have been spatially interpolated to the subbasins. The potential evapotranspiration is derived from long-term mean daily evapotranspiration values which have been corrected by the effect of actual temperature [25]. The model has been calibrated for the 1996–2007 period.

For application of the spatially-distributed REW-model, the catchment of gauge Bollendorf is subdivided into nine Representative Elementary Watersheds (REWs) and the one of gauge Trier into 85 entities. The REWs represent irregularly-shaped modelling elements for which relevant hydrological processes are identified and modelled. The unsaturated zone is further broken down into subunits based on land-use and soil type. The main governing equations are upscaled point-scale conservation equations. Physical governing equations, such as the Richards equation, are also used to describe vertical moisture diffusion through the unsaturated soil column. Hourly time series of precipitation, temperature and relative air humidity interpolated toward REWs force the model at 15 min time steps. The model was calibrated for the 1996–2011 period.

Recurrent neural networks (RNN) [29] are inspired by the information processing in biological neural networks. The processing elements represent neurons which are linked by connections. This connection topology contains feedback loops, and therefore previously received input signals are still represented in the net (*i.e.*, implement some memory) and influence its output. Mathematically, a RNN is a dynamical system akin to a system of differential equations. RNNs are able to model the input-output behavior of dynamical systems which have states evolving over time. The same temperature and precipitation time series used for the HBV subbasins are applied as input time series to the RNN. Echo State Networks ESN [28,29] is a method designed for constructing and training large RNNs. ESNs are built in a special way, thereby ensuring the so called "Echo State property", and

then only the connections to the output nodes are computed in the training phase. This is done using regression, which is far more simple and stable than gradient-based optimization. This simplicity enables the training of large networks including thousands of internal nodes. These large nets have the capacity of learning high-dimensional dynamics of complex systems. Two ESN have been trained for each gauge for the 2000–2012 period and two flow situations: one with focus on low to mean flow and the second with focus on mean to high flows. The two outputs are combined on the basis of the specific situation reproduced by the ESNs. Each ESN had one output node (*i.e.*, the predicted flow), 2000 internal nodes, and eight (gauge Bollendorf) and 40 (gauge Trier) input nodes, *i.e.*, temperature and precipitation for each HBV subbasin to be considered.

The chief aim of this study is to demonstrate PU assessment using copulas rather than addressing hydrological calibration and validation strategies. Hence, model simulations are used "as is", despite inconsistencies between different calibration and validation periods. The uncertainty processor was applied to the 1 October 2000 to 30 September 2015 period.

2.3. Copula Theory

A copula is a function that describes the dependence structure of correlated random variables independently from marginal distributions. Due to their flexibility in modelling different dependence structures, copulas are increasingly applied in multivariate analysis for flood frequency analysis, drought analysis, dam risk assessment and rainfall processing [30–33]. In hydrological forecasting, Madadgar *et al.* [18] and Madadgar and Moradkhani [22] applied copulas for statistical post-processing of hydrological ensemble forecasts.

For sake of brevity, we give only a short overview about copulas here. For a more detailed description of the theory, the reader is referred to [34–36].

After Sklar's theorem [37] a joint multivariate probability distribution of n correlated variates X_1, X_2, X_3, $\ldots$, X_n can be described through the copula function C. The copula function is a multivariate distribution function of an n-dimensional random vector on the unit cube, *i.e.*, $C: [0,1]^n \rightarrow [0,1]$, and is strictly invariant under the condition of monotonously increasing transformations of X_1, X_2, X_3, $\ldots$, X_n:

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, ..x_n) = C\left[F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_n}(x_n)\right] = C(u_1, u_2, ..., u_n) \tag{1}$$

where $u_1, u_2, \ldots, u_n \in [0,1]$ are uniformly distributed random realizations of the variates defined as $u_1 = F_{X_1}(x_1)$, $u_2 = F_{X_2}(x_2)$, $\ldots$, $u_n = F_{X_n}(x_n)$.

The joint multivariate density function is expressed via the copula density c as follows:

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, ..x_n) = \frac{\partial^n C(u_1, u_2, ..., u_n)}{\partial u_1 \partial u_2 ... \partial u_n} \prod_{i=1}^{n} f_{X_i}(x_i) = c(u_1, u_2, ..., u_n) \prod_{i=1}^{n} f_{X_i}(x_i) \tag{2}$$

2.4. Pair-Copula Construction

For the bivariate case, which models the dependency of two random variables, a large number of copulas are available [34,35]. By contrast the number of available copula functions for the multivariate case is rather limited and the applicable elliptical (Gaussian copulas, Student t copulas) and exchangeable Archimedian copulas lack the flexibility to model dependency structures of multiple heterogeneous random variables. To overcome this problem, the concept of Vine copulas has been developed [38–41]. Vine Copulas are a flexible graphical model to build multivariate copulas from a cascade of bivariate or pair-copulas. The advantage of this pair-copula construction is the decomposition of the multivariate probability density into several bivariate copulas, which can be selected independently. Vine copulas have been originally applied in the area of risk-modelling in the financial sector and more recently also in hydrology [42–44].

As there are a significant number of possible pair-copula constructions for high-dimensional distributions (e.g., 240 different constructions for a five-dimensional density [45]), Bedford and Cooke [39,40] introduced the graphical model, the so called Regular Vines, which help to organize possible combinations. In this application, we concentrate on the special case of C(anonical)-Vines [45], which we consider the most suitable pair-copula construction for the estimation of the PU of hydrological multi-model ensembles. All conditional random variables can be directly linked to the random variable of interest, the observation (random variable 1 in Figure 2).

Figure 2 shows the graphical representation of a four-dimensional C-Vine copula. It consists of a number of random variables $n-1$ trees T, Tree T_j has $n+1-j$ nodes and $n-j$ edges. In total, the decomposition is defined by $n(n-1)/2$ edges and each edge is represented by a pair-copula density. The labels of the edges represent subscript of the copula density, e.g., 13|2 denotes the copula density $c_{13|2}(\cdot)$ [39,45].
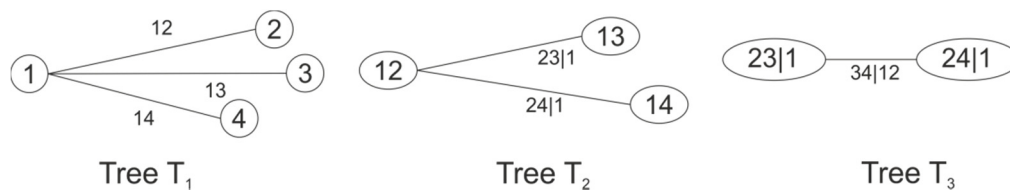

Figure 2. Graphical representation of a 4-dimensional C-Vine copula [45].

The n-dimensional density of the C-Vine copula is defined as:

$$f(x_1, \ldots x_n) = \prod_{k=1}^{n} f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,j+i|1,\ldots,j-1}\left[F(x_j|x_1,\ldots,x_{j-1}), F(x_{j+i}|x_1,\ldots,x_{j-1})\right] \quad (3)$$

with j sweeping the trees and i sweeping the edges in each tree.

The conditional distribution functions for each node in a tree are determined by sequentially applying the following equation over the previous trees:

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}\left[F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})\right]}{\partial F(v_j|\mathbf{v}_{-j})} \quad (4)$$

with the m-dimensional vector $\mathbf{v}$, v_j an arbitrary element of $\mathbf{v}$ and $\mathbf{v}_{-j}$ the $(m-1)$-dimensional vector $\mathbf{v}$ excluding element v_j.

When applying C-Vines, the PU of the random variable observation (random variable X_4 in Figure 2) conditional on other variables can be solved analytically by means of Equation (4). In this case, the key variable has no direct connection to the remaining variables in the most important first tree, and the dependence is only modelled indirectly in the downstream trees. Hence, it is recommended to locate the key variable at the root (variable X_1 in Figure 2) of the C-Vine copula [45]. The conditional copula density of the root variable conditional on the other variables is:

$$c_{1|2,3,\ldots,n}(u_1|u_2,u_3,\ldots,u_n) = \frac{c_{1,2,3,\ldots,n}(u_1,u_2,u_3,\ldots,u_n)}{c_{2,3,\ldots,n}(u_2,u_3,\ldots,u_n)} \quad (5)$$

As the pair-copula only "approximates" the full copula density, it is difficult to derive the lower-dimensional copula density in the denominator analytical. In this application, we calculated the copula density in the denominator by integrating $c_{1,2,\ldots,n}$ numerically over the first variable between 0 and 1 while fixing the remaining ones.

The PU of the observation y, conditional on multiple hydrological predictions $\hat{y}$, can be finally expressed via the numerical integration of Equation (5) as:

$$F\left(y\middle|\hat{y}_1, \hat{y}_2, ..., \hat{y}_n\right) = \int_0^{F(y)} c_{1|2,3,...n}\left(u_1\middle|u_2 = F\left(\hat{y}_1\right), u_3 = F\left(\hat{y}_2\right), ..., u_n = F\left(\hat{y}_n\right)\right) du_1 \qquad (6)$$

The selection of the pair-copula families and the parameter estimation of the C-Vine copulas is done stepwise [46]. In the selection process, the appropriate pair-copula functions are selected sequentially tree by tree as the conditional pairs of the higher order trees depend on the previous trees. The copula functions are chosen via Akaike's Information Criteria (AIC) [47] and the canonical maximum likelihood method [48] for parameter estimation among the following options [34,35,46]: Gaussian, Student t, Gumbel, Joe, BB1, BB6, BB7, BB8 copula as well as the corresponding survival copulas (rotated by 180 degrees) of the Clayton, Gumbel, Joe, BB1, BB6, BB7, BB8. The canonical maximum likelihood method estimates the parameters of the copula function independently of the uncertain estimation of the marginal distributions, as it uses empirical probabilities to transform the sample data into the uniform variates. After selecting the copula families and estimating their parameters for all edges sequentially, the parameters of the whole C-Vine structure are optimized simultaneously by maximizing the log-likelihood of the C-Vine copula [46]. The goodness-of-fit of the pair-copulas are analyzed graphically by comparing the bivariate sample values with simulated values from the estimated copula in the copula space and for the copulas of the first tree also in the original space. Moreover, the empirical and theoretical λ-functions [49] defined as $\lambda = v - K(v)$, with $K(v)$ Kendall's distribution of the random variable $C(u_1, u_2)$ are also compared. In this application, the C-Vine choice and parameter estimation was performed with the R-package CDVine [46].

2.5. Marginal Distributions and Normal Quantile Transform

In most of the existing approaches for PU estimation, meta-Gaussian models are applied to describe the joint distribution between the transformed real (observed) and simulated values of the variable of interest. Simulated and observed values are transformed to the Normal space by means of the Box-Cox transformation [50], log-sinh transformation [51] and the non-parametric Normal Quantile Transform (NQT) [52–54], which, however, has limitations in extrapolating the relationship between the values in the Normal and real space beyond the range of the historically observed values. The non-parametric NQT relationship between the values in the real space and the values in the Normal space are calculated from the empirical probabilities and the corresponding quantiles of a standard Normal distribution. A detailed description of the application of the NQT can be found in [2]. Parametric models are needed at the tails of the NQT to extrapolate values beyond the range of available observations. Several methods have been developed to handle this problem [11,55].

In the application described here, we use parametric probability distribution functions as marginal distributions $F(y)$ for the application of copulas according to Equation (6), and to transform the observed and predicted values y to the Normal space η when applying other PU estimation methods for inter-comparison of results (see Section 2.6):

$$\eta = \Phi^{-1}\left(F\left(y\right)\right) \qquad (7)$$

with Φ^{-1} the inverse cumulative distribution function of the standard Normal distribution.

Estimating parametric probability distributions of daily flow rates and/or water levels is often difficult, as the values arise from completely different meteorological and hydrological situations (e.g., low flows, high flows) and can in most cases not be described via one single probability distribution.

For this reason, a mixture of probability distributions has been developed to better represent the distributions of heterogeneous data [56–58]. Solari and Losada [58] used a mixture of probability distributions with a truncated central distribution and two generalized Pareto distributions for the

maximum and minimum regimes of mean daily flow time series. The latter approach was used here to estimate the marginal distributions of the random variables.

By considering two thresholds u_1 and u_2, the probability density function (pdf) and the cumulative distribution function (cdf) of the mixed probability distribution are defined as:

$$f(y) = \begin{cases} f_m(y) \, F_c(u_1) & y < u_1 \\ f_c(y) & u_1 \leqslant y \leqslant u_2 \\ f_M(y) \, [1 - F_c(u_2)] & y > u_2 \end{cases} \tag{8}$$

$$F(y) = \begin{cases} F_m(y) \, F_c(u_1) & y < u_1 \\ F_c(y) & u_1 \leqslant y \leqslant u_2 \\ F_c(u_2) + [1 - F_c(u_2)] \, F_M(y) & y > u_2 \end{cases} \tag{9}$$

The multiplication of the lower distribution $f_m(y)$ with $F_c(u_1)$ and the upper distribution $f_M(y)$ with $[1 - F_c(u_2)]$ is required to ensure the area under the probability density graph to equal unity.

For the lower distribution a minima Generalized Pareto (GP) distribution and for the upper distribution a GP distribution is used. The type of the central distribution function can be selected independently of the upper and lower distribution as long as it is defined in the range u_1 to u_2.

The pdf of the minima GP distribution is a mirrored ($y = -y$) GP distribution density (defined as maxima GP distribution in the following) and the pdf and cdf are therefore defined as:

$$f_m(y) = \frac{1}{a_1} \left[1 - \frac{k_1}{a_1} (u_1 - y) \right]^{\frac{1}{k_1} - 1} \tag{10}$$

$$F_m(y) = \left[1 - \frac{k_1}{a_1} (u_1 - y) \right]^{\frac{1}{k_1}} \tag{11}$$

The minima GP distribution has an upper boundary u_1 for negative k: $y \leqslant u_1$; $k_1 < 0$ and a lower and upper boundary for positive k: $u_1 - a_1/k_1 \leqslant y \leqslant u_1$; $k_1 > 0$.

The pdf and cdf of the maxima GP are defined as:

$$f_M(y) = \frac{1}{a_2} \left[1 - \frac{k_2}{a_2} (y - u_2) \right]^{\frac{1}{k_2} - 1} \tag{12}$$

$$F_M(y) = 1 - \left[1 - \frac{k_2}{a_2} (y - u_2) \right]^{\frac{1}{k_2}} \tag{13}$$

The maxima GP distribution has a lower boundary u_2 for negative k: $u_2 \leqslant y$; $k_2 < 0$ and a lower and upper boundary for positive k: $u_2 \leqslant y \leqslant u_2 + a_2/k_2$; $k_2 > 0$.

In the parameter estimation process of the mixture probability distribution, one needs to make sure that no discontinuities occur in the transition of the pdfs at the thresholds u_1 and u_2. Using the conditions $f_c(y = u_1) = f_m(y = u_1) \, F_c(u_1)$ and $f_c(y = u_2) = f_M(y = u_2) \, [1 - F_c(u_2)]$ the parameters a_1 and a_2 of the generalized Pareto distributions are obtained as:

$$a_1 = \frac{F_c(u_1)}{f_c(u_1)}; \quad a_2 = \frac{1 - F_c(u_2)}{f_c(u_2)} \tag{14}$$

By defining a lower b_l and upper boundary b_u of the mixture probability distribution as an additional condition, the shape parameters k_1 and k_2 of the GP distribution result equal to:

$$k_1 = \frac{a_1}{u_1 - b_l}; \quad k_2 = \frac{a_2}{b_u - u_2} \tag{15}$$

The remaining parameters of the mixture probability distribution left to estimate are the central distribution and the thresholds u_1, u_2. All parameters are estimated simultaneously using maximum likelihood estimation. Solari and Losada [58] used a two-parameter log-Normal distribution as central distribution function f_c. In this application, different types of central distributions (generalized Extreme Value, Logistic, generalized Logistic, Pearson3, log-Pearson3, Frechet, Gamma, Normal, log-Normal, Weibull, log-Weibull, Gumbel, Exponential, and generalized Pareto distribution [59]) were applied and the type of distribution with the best fit based on the AIC was selected.

2.6. Alternative Methods applied for PU Estimation

The results of the COpula uncertainty Processor (COP) defined in Equation (6) are compared against the results of three frequently used methods: Quantile Regression (QR) [13,14], Bayesian Model Averaging (BMA) [15–18] and Multivariate Truncated Normal (MTN) distribution as e.g., applied in the Model Conditional Processor [11]. Only a short overview of the theory of the methods is given here. For further details, we refer to the cited publications. When applying these methods, the observed and simulated values y are first transformed to the Normal variable η (model predictions are indicated by the hat symbol) and the estimated quantiles of the predictive uncertainty are back-transformed to the real space by the inverse transform. Here, the values are transformed to the Normal space by means of the same parametric mixture probability distributions used as marginal distributions in the COP via Equation (7). Hence, differences in results are attributable exclusively to the specific PU method and do not depend on the marginal distribution or the transformation method.

As done with QR, we fit separate multiple linear regression relationships for each quantile τ via optimized linear programming [60] using the model predictions as predictors. To approximate the predictive distribution, multiple quantiles need to be estimated using:

$$\eta_\tau | \hat{\eta}_1, \hat{\eta}_2 \ldots, \hat{\eta}_M = a_\tau + \sum_{i=1}^{M} b_{\tau,i} \hat{\eta}_i \tag{16}$$

with a_τ, $b_{\tau,\iota}$ the parameters of the multiple linear regression relationship and M the number of models. The R-package quantreg [61] was used for parameter estimation.

The BMA predictive distribution is a linear combination of conditional probability distributions of observation, conditional on model predictions and given that the prediction is the best possible in the ensemble. Assuming Normal conditional distributions, the BMA predictive uncertainty is:

$$f\left(\eta | \hat{\eta}_1, \hat{\eta}_2 \ldots, \hat{\eta}_M\right) = \sum_{i=1}^{M} w_i N\left(\eta | a_i + b_i \hat{\eta}_i, \sigma_i^2\right) \tag{17}$$

where w_i is the posterior probability of prediction i being the best, and M the number of models. The weights w_i are non-negative and add up to unity [15]. The conditional Normal distribution of the individual model prediction is centered on a linear dependency for the model prediction. Parameters a_i and b_i act as bias correction factors. The weights w_i and standard deviations σ_i are estimated from past model performance. We applied a two-step parameter estimation procedure: (1) maximize log-likelihood with the Expectation Maximum (EM) algorithm to estimate weights and standard deviations and (2) minimize the Continuous Ranked Probability Score (CRPS) by applying Equation (23) to optimize the estimation of the standard deviations. This procedure is implemented in the R-package ensemble BMA [62].

To consider error heteroscedasticity in PU estimation [2,11], MTN distributions can be used instead of the simple multivariate Normal distribution. Here, a combination of two MTN distributions is used. The split of the multivariate Normal space into two parts is obtained by identifying an *M*-dimensional hyperplane:

$$H_p = \sum_{i=1}^{M} \hat{\eta}_i = M \cdot a \tag{18}$$

with *M* the number of prediction models. The value of *a* is identified by minimizing the predictive variance of the upper sample. The predictive distribution for the sample above the truncation hyperplane is defined as:

$$f\left(\eta \,|\, \hat{\eta}_k = \hat{\eta}_k^*, H_P^* > M \cdot a\right) = N\left(\mu_{\eta|\hat{\eta}_k=\hat{\eta}_k^*, H_P^*>M \cdot a}, \sigma^2_{\eta|\hat{\eta}_k=\hat{\eta}_k^*, H_P^*>M \cdot a}\right) \tag{19}$$

with the predictive mean

$$\mu_{\eta|\hat{\eta}_k=\hat{\eta}_k^*, H_P^*>M \cdot a} = \mu + \sum_{\eta\hat{\eta}} \Sigma^{-1}_{\hat{\eta}\hat{\eta}}\left(\hat{\eta}_k^* - \hat{\mu}\right) \tag{20}$$

and predictive variance

$$\sigma^2_{\eta|\hat{\eta}_k=\hat{\eta}_k^*, H_P^*>M \cdot a} = \Sigma_{\eta\eta} - \Sigma_{\eta\hat{\eta}}\Sigma^{-1}_{\hat{\eta}\hat{\eta}}\Sigma^T_{\eta\hat{\eta}} \tag{21}$$

The model parameters mean, variance and covariance matrices are estimated from the data of the upper sample. The predictive distribution of the lower sample looks similar but is defined for the values of the sample below the truncation hyperplane $H_P^* \leqslant M \cdot a$. For more details see [11].

2.7. Verification

We base the verification of the predictive distributions estimated with the methods described above on the paradigm of *maximizing the sharpness of the predictive distributions subject to calibration* [63]. Calibration refers to the statistical consistency between the predictive distribution and observations. The percentage levels of prediction intervals should be equal to the relative frequency of the observations to lie within the corresponding quantiles of the predictive distribution. Another common term for calibration in the present context is "reliability". Sharpness refers to the concentration of the predictive distribution. The more concentrated ("narrow" prediction intervals) the predictive distribution, the sharper the prediction.

Reliability is assessed via the probability integral transform (PIT) [63]. From predictive distribution the non-exceedance probability is estimated for each value observed. These probabilities should have a uniform distribution, *i.e.*, the observations should look like random samples from the predictive distribution. When deriving the relative frequencies of the non-exceedance probabilities for defined 10% quantile intervals, one obtains the PIT-histogram. A flat histogram indicates good calibration, whereas under-dispersion is indicated by a ∪-shape and overdispersion by a ∩-shape, respectively. Sharpness is assessed by the distribution within the 5%–95% credibility interval of each verification day. The distribution is illustrated by a modified Box-Whisker-Plot: the box represents the 25%–75% quantile, the median is the band inside the box and the whiskers represent the 5% and the 95% quantile. Outliers are not shown. Additionally, the predictive distribution is verified by the probabilistic verification measures Brier Skill Score (BSS) and the Continuous Ranked Probability Skill Score (CRPSS), two combined measures of reliability, sharpness and accuracy.

The Brier Score (BS) is a common verification measure of binary events (yes/no). It expresses the mean squared error of the probability forecast p_i, considering that observation o_i is 1 if the event occurs and 0 in its absence [21]:

$$BS = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2 \tag{22}$$

with n the number of verification days. In this application, the event is defined as the exceedance *vs.* non-exceedance of flow rate thresholds ranging from the 5% quantile to the 95% quantile of observed daily mean flow values.

The Continuous Ranked Probability Score (CRPS) verifies the full continuous distribution and is the squared difference between the predictive distribution F_i^f and the distribution of the observation F_i^o [21]:

$$CRPS = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left[F_i^f(y) - F_i^o(y)\right]^2 dy \tag{23}$$

whereas the probability distribution of the observation is a cumulative-probability step function that jumps from 0 to 1 at the point where the variable y equals the observation.

The corresponding skill scores $CRPSS = 1 - CRPS_{FOR}/CRPS_{REF}$ and $BSS = 1 - BS_{FOR}/BS_{REF}$ are calculated from the $CRPS_{FOR}$ and BS_{FOR} of the predictive distribution and the $CRPS_{REF}$ and BS_{REF} of the reference forecast, for which the climatological forecast, *i.e.*, the observed daily mean flow of the period from 1 October 2000 to 30 September 2015 is used. Due to the strong seasonality of the flow at the two gauges (see Figure 1) the $CRPSS$ and BSS are calculated for each month separately and the climatological forecast is calculated only from the observed values of the corresponding month excluding the current month of verification. A perfect forecast has a skill score of one, negative skill score values indicate that the climatological forecast is superior to the predictive uncertainty derived from the model predictions.

The accuracy of the expected value of the predictive distribution as well as the goodness-of-fit of the deterministic model predictions is verified with the modified Kling-Gupta efficiency KGE' and its respective components [64]:

$$KGE' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{24}$$

$$\rho = \frac{Cov(y,\hat{y})}{\sigma_{\hat{y}}\sigma_y} \tag{25}$$

$$\beta = \frac{\mu_{\hat{y}}}{\mu_y} \tag{26}$$

$$\gamma = \frac{CV_{\hat{y}}}{CV_y} = \frac{\sigma_{\hat{y}}/\mu_{\hat{y}}}{\sigma_y/\mu_y} \tag{27}$$

where ρ is the Pearson correlation between the observation and prediction, β is the bias ratio, γ is the variability ratio which uses the ratio between the coefficient of variation of the prediction and the observation. All components have their optimum at unity and the KGE' gives the lower limit of the three components [64].

The expected value of the predictive distribution is calculated as the mean of 1%-quantiles of the predictive distribution.

3. Results and Discussion

We applied the COP introduced here and the other uncertainty processors to estimate the PU of the observed daily mean flow values conditional on model predictions at two gauges located in the Moselle river basin for the 1 October 2000 to 30 September 2015 analysis period. To ensure that no flow values are simultaneously used for parameter estimation and validation of the predictive distribution, we used a Leave-One-Out Cross Validation (LOOCV) procedure. For each validation year, the parameters of the uncertainty processors were estimated from the values of the study period, by excluding the values of the validation year. In total, the parameters of each uncertainty processor were updated once per validation year.

3.1. Model Predictions

The goodness-of fit of the three model predictions is reported in Table 2 for the whole 1 October 2000 to 30 September 2015 study period. All models are in better agreement with observations at gauge Trier, which has the seven-fold upstream catchment area of gauge Bollendorf, based on the goodness-of-fit measure modified Kling-Gupta Efficiency KGE′ and its components. From the values in the table, it is evident that the KGE′ provides the lower limit of its respective components. At both gauges, the predictions of the HBV model show significant bias, whereas the ESN model has difficulties in capturing the variability of the observations at both gauges. The REW predictions show a significant bias only at the gauge Bollendorf. Concentrating on the linear relationship between observations and predictions in terms of cross-correlations, the HBV model shows the best results at both gauges. Overall, no model outperforms the others at both gauges and for all measures.

Table 2. Goodness-of-Fit (GoF) measure KGE′ and its components for the two hydrological model predictions HBV, REW and the prediction with the echo state network model ESN for the period 1 October 2000 to 30 September 2015 at the gauges Bollendorf and Trier.

GoF Measure	Bollendorf			Trier		
	HBV	**ESN**	**REW**	**HBV**	**ESN**	**REW**
KGE'	0.85	0.78	0.75	0.83	0.87	0.85
Correlation ρ	0.94	0.89	0.89	0.95	0.95	0.89
Bias-Ratio β	1.10	0.99	1.22	0.84	1.03	1.01
CV-Ratio γ	0.91	0.80	1.05	0.95	0.88	1.10

3.2. Estimation of the Marginal Distributions

In a first step of the PU estimation using the COP (Equation (6)), the marginal distributions of the random variables need to be estimated. Mixture probability distributions (Equations (8) and (9)) have been fitted to the observed mean daily flow values as well as to the HBV, REW and ESN predictions at the two gauges.

In the parameter estimation process, a flow rate 0 m^3/s was used as lower and three times the maximum rate as the upper boundary. For the distribution of the observed values, the maxima shown in Table 1 and for the distributions of the predictions the maximum values of the simulated time series 1 October 2000 to 30 September 2015 were used. The types of the central distributions have been selected based on the AIC, using the maximum likelihood method for parameter estimation.

Figure 3 shows the cumulative distribution function (cdf) plots of the mixture distributions fitted to the observed daily mean flow at gauges Bollendorf and Trier and the two thresholds u_1 and u_2. At Bollendorf, the identified central distribution of the mixture distribution was the log-Pearson3 and, at Trier, the generalized Pareto distribution using all data of the 1 October 2000 to 30 September 2015 analysis period. The selection of the type of the central distribution is relatively stable. When using the LOOCV subsample sets at gauge Bollendorf, the same distribution (log-Pearson3) was identified 10 times out of 15 subsamples and, at gauge Trier, the same distribution (generalized Pareto distribution) was identified for all subsamples. The variability of the fitted distributions and the two thresholds u_1 and u_2 for the different subsamples in the LOOCV are shown in Figure 3 as dotted lines.

At gauge Bollendorf, several values are above the theoretical distribution at the upper tail. This is due to the fact that the highest observed flood in the last 56 years at gauge Bollendorf falls within the 15-year study period. The daily flow rates of this flood have therefore a lower probability of occurrence than the empirical probabilities shown in Figure 3. The overall fit of the modeled distributions is satisfactory on visual inspection. For the mixture distributions of the individual model predictions, other central distributions have been identified in the selection process.
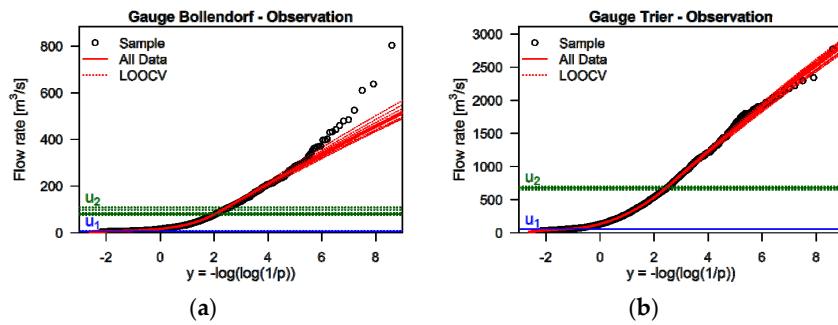
Figure 3. Cumulative distribution function (cdf) plots of the mixture distributions (**red**) fitted to the observed daily mean flow at (**a**) Gauge Bollendorf; and (**b**) Gauge Trier and the two thresholds u_1 (**blue**) and u_2 (**green**) using all data (solid) and Leave-One-Out Cross Validation (LOOCV) subsample sets (dotted) for parameter estimation. The probabilities on the *x*-axis of the cdf are in a log-log scale.

3.3. C-Vine Copula Selection and Estimation

In the next step of the COP estimation, the C-Vine structure needs to be defined. As the central variate 1 in Figure 2, we chose the "observations" because it is directly linked to the other variables in the first tree. The other random variables in Figure 2 are HBV (2); ESN (3); and REW (4). Other combinations of variates 2–4 would lead to slightly different but overall similar results. The bivariate copula families have been selected based on the AIC using the canonical maximum likelihood method for parameter estimation.

The scatter plots in the second column of Figure 4 show that all bivariate random samples have a different dependence structure but all of them indicate an upper tail dependence, implying that the two variates behave similarly at the upper tail of the distribution. Hence, we use only bivariate copulas, which allow modeling upper tail dependence. The frequently used Clayton and Frank copulas are not selected because of their poor capability to represent upper tail dependence.
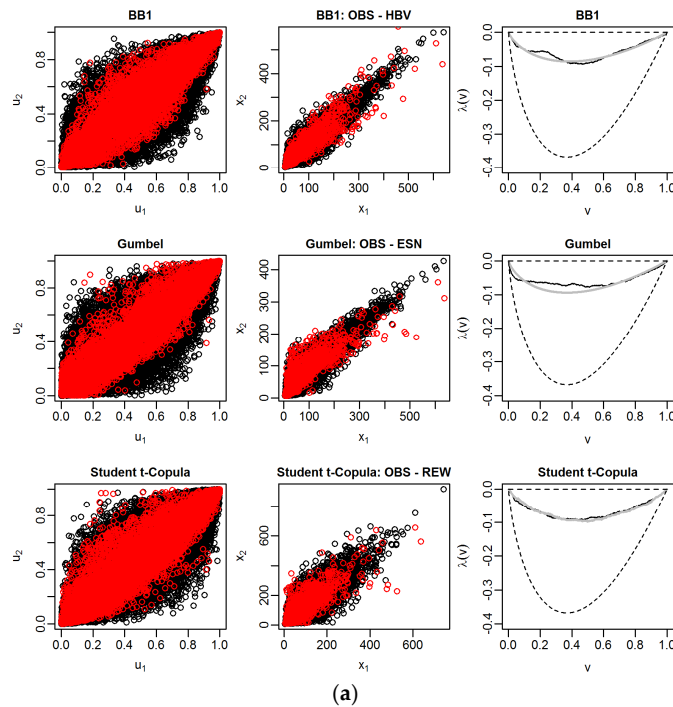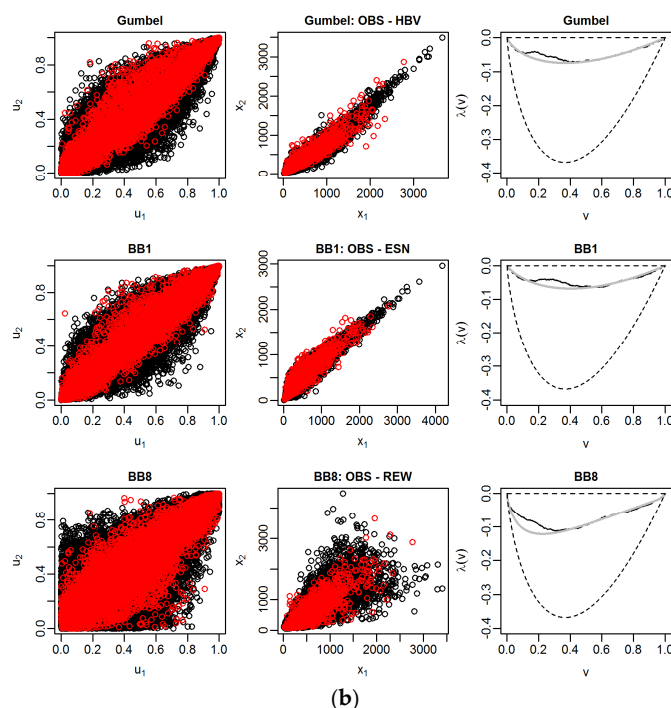

(**a**)

Figure 4. *Cont.*

(b)

Figure 4. Estimated bivariate copulas of the first tree of the C-Vine for the bivariate samples observation and HBV prediction (first row); observation and ESN prediction (second row); observation and REW prediction (third row) of (**a**) gauge Bollendorf; and (**b**) gauge Trier using all data of the analysis period. Shown are the scatter plots of the bivariate sample (**red**) and a random sample (10 times the size of the bivariate sample) from the fitted copula (**black**) in the copula space (first column) and back-transformed to the real space (second column). In the third column, the empirical λ-functions (**black**) and the theoretical λ-functions (**grey**) are shown. The dashed lines are bounds corresponding to independence and comonotonicity ($\lambda = 0$), respectively.

Figure 4 illustrates the goodness-of-fit of the estimated bivariate copulas of the most important first tree of the C-Vine copula at both gauges using all data of the 1 October 2000 to 30 September 2015 analysis period.

In the first column, the scatter plots of random pairs chosen from the selected pair-copulas with the sample values in the copula space are compared, while in the second one the back-transformed values using marginal distributions are compared in the real space. The size of the respective random samples is 10 times the size of the bivariate sample. Comparing simulated and sample values in the two scatter plots, we can conclude that the fitted copulas are able to capture the dependence structure of the bivariate samples. In the third column the empirical λ-functions (black) of the bivariate samples and the theoretical λ-functions (grey) of the estimated pair-copulas are shown. The dashed lines below and above are bounds corresponding to independence (lower line) and comonotonicity ($\lambda = 0$), *i.e.*, perfect positive dependence, respectively. As the empirical λ-functions are near the line of co-monotonicity, there is a strong dependence between all considered combinations of the random variables. Overall, the fitted copulas show good agreement with empirical values. We note that for nearly all pair-copulas estimated at both gauges, there is a deviation between theoretical and empirical functions when the random variable v = $C(u_1, u_2)$ takes on values between 0.1 and 0.3. No theoretical copula function exists, which is able to capture the irregularity in the λ-function of the bivariate sample data present at both gauges.

Table 3 shows the identified pair-copula families using all data and the LOOCV subsample sets for the gauge Trier. It is evident that the identification of the pair-copulas is relatively stable as the same copulas are identified for nearly all subsamples. In the LOOCV procedure, these copula families are used to estimate the PU.

Table 3. Pair-copula families of the C-Vine identified using all data and LOOCV (validation year as identifier) at gauge Trier.

LOOCV	Tree 1			Tree 2		Tree 3
Sample	$C_{Obs,HBV}$	$C_{Obs,ESN}$	$C_{Obs,REW}$	$C_{HBV,ESN\mid Obs}$	$C_{HBV,REW\mid Obs}$	$C_{ESN,REW\mid Obs,HBV}$
All Data	Gumbel	BB1	BB8	Student t	Student t	Student t
2001, 2014	Gumbel	BB1	BB8	Student t	Student t	Independence
2002, 2004	BB1	Student t	BB8	Student t	Student t	Independence
2003, 2005, 2006, 2007, 2008, 2009, 2010, 2013, 2015	Gumbel	BB1	BB8	Student t	Student t	Student t
2011	Gumbel	Student t	BB8	Student t	Student t	Independence
2012	BB1	BB1	BB8	Student t	Student t	Independence

3.4. Verification

The quality of the estimated predictive distributions using the COP (Equation (6)) and results of BMA, QR and MTN distribution (Equations (16), (17) and (19)) are verified and compared below. Based on the verification metrics presented before, the measures sharpness, reliability and accuracy of the estimated predictive distributions are verified and compared for the whole study period. All verification results are based on the estimated predictive distributions from the LOOCV strategy. Hence, no values are used simultaneously for parameter estimation and verification.

In Figure 5, the reliability of the estimated predictive distributions is depicted as Probability Integral Transform (PIT)-histograms. For the predictive distribution estimated with the COP, the number of observations in the lower quantile ranges (10%–40%) are slightly below and the number of observation in the lowest quantile range (0%–10%) are above the optimal value of 10% at both gauges. The shape of the COP PIT-histogram at both gauges could be an indication of a small negative bias. The shapes of the PIT-histograms of BMA and MTN indicate over-dispersion at both gauges, which suggests that the estimated variance of the distributions is too large. The highest reliability indicated by the flat shape of the histogram is that of the QR approach, which is reasonable since the parameters of QR quantiles are optimized based on reliability. Moreover, in QR, the number of fitted parameters is considerably higher than that of the other methods. For each quantile, four parameters need to be estimated. This means that, in total, 36 parameters have been estimated for the nine quantiles 10%–90% used for reliability verification. Despite small overall differences, we conclude that all estimated predictive distributions can be considered reliable with outperformance of COP and QR.
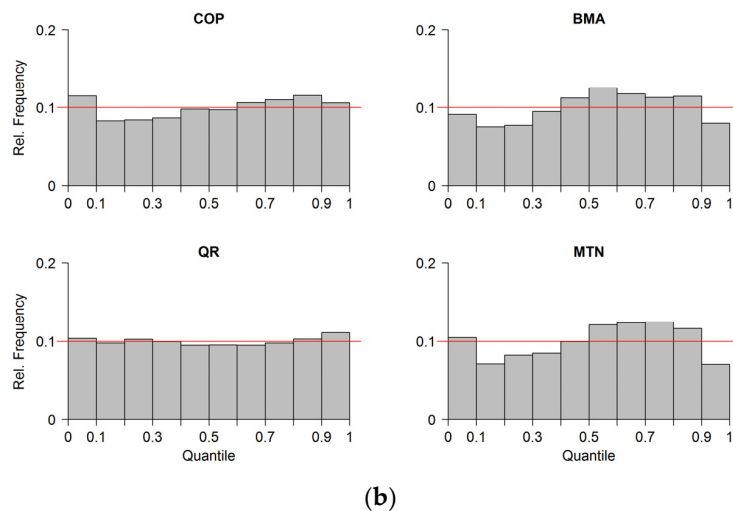

(a)

Figure 5. *Cont.*

(b)
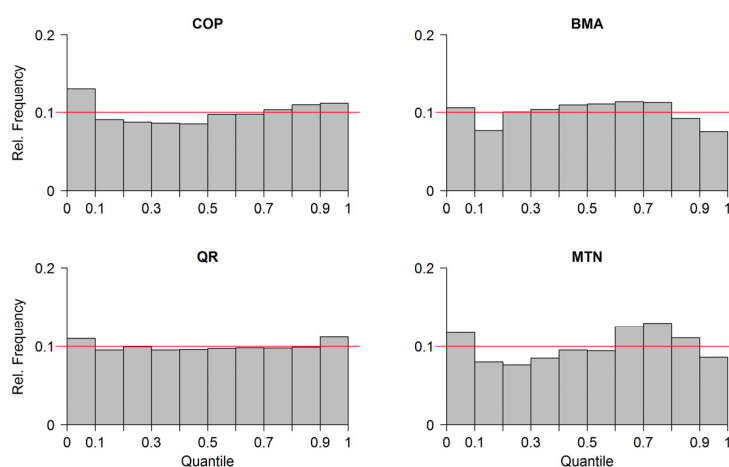
Figure 5. PIT-histograms (**a**) gauge Bollendorf; (**b**) gauge Trier. **Upper left** COpula uncertainty Processor (COP), **lower left** quantile regression (QR) **upper right** Bayesian Model Averaging (BMA), **lower right** Multivariate Truncated Normal (MTN) distribution.

Verification of reliability as well as calibration is only meaningful in combination with the assessment of sharpness, which is an attribute of the prediction alone. For sharpness assessment of the predictive distributions, the width of the 5%–95% quantile range is calculated for each verification day. The smaller the widths of the quantile range the sharper and more adherent the predictive distribution is. Figure 6 shows important quantiles of the distribution of the widths as modified box plots.

The upper part of Figure 6 illustrates the coverage, which measures the proportion of observations that fall within the 5%–95% quantile range. In principle, 90% of observations should fall within this interval, which is the case in all applied methods. QR and COP provide sharper predictions than MTN and BMA. Other quantile ranges analyzed (10%–90%, 25%–75%, not shown here) show similar tendencies.
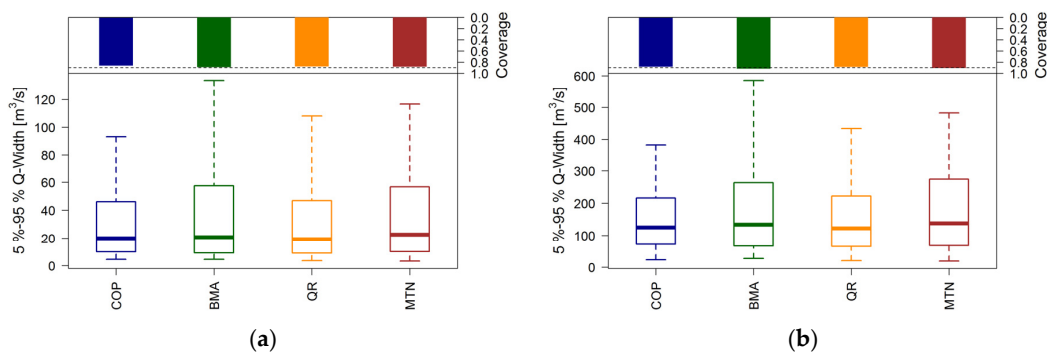

Figure 6. Distributions of the 5%–95% quantile widths are shown as modified boxplots at (**a**) gauge Bollendorf; (**b**) gauge Trier. The box represents the 25% to 75% quantile, the median is the band inside the box and the whiskers represent the 5% and the 95% quantile. Outliers are not shown. In the upper part of the figure, bars represent the "coverage", *i.e.*, the percentage of observations falling in the 5%–95% quantile range.

To assess the predictive performance of the estimated distributions for different runoff situations (low flow, mean flow, high flow), we calculate the Brier Skill score for the non-exceedance probabilities of different flow rate thresholds, ranging from the 5%–95% quantile of the observations. Figure 7 shows the BSS values at the two gauges.

At both gauges, all methods show similar skill in terms of BSS for all flow rate thresholds. At gauge Bollendorf, QR and MTN show a higher skill for higher quantiles and a lower skill for smaller quantiles than COP and BMA. Overall, the predictive performance of all methods is comparable at both gauges. We note a local maximum at the 25% quantile of all methods at both gauges. For larger quantiles, the BSS decreases and then rises again with threshold values higher than the 50% quantile.
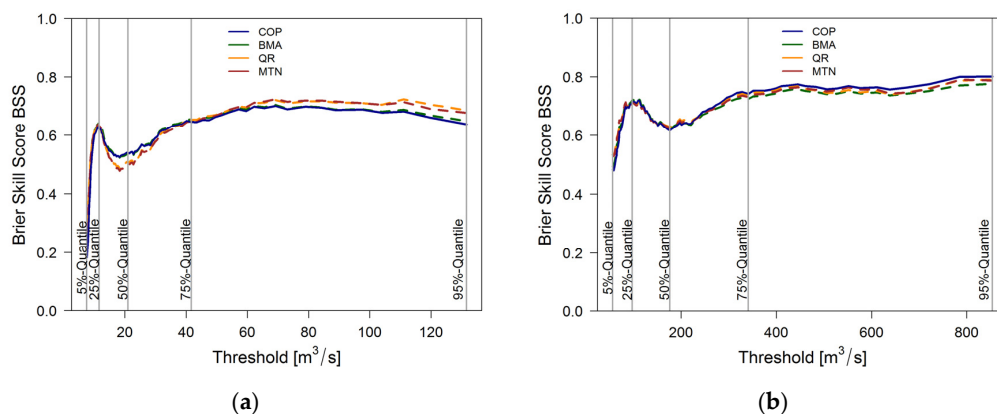

Figure 7. Brier Skill Score for different flow rate thresholds (**a**) gauge Bollendorf; (**b**) gauge Trier. COpula uncertainty Processor (COP) blue line, Bayesian Model Averaging (BMA) green line, quantile regression (QR) orange line and Multivariate Truncated Normal (MTN) distribution brown line.

This effect could correspond to the irregularities observed in the empirical λ-function of the bivariate samples at both gauges in Section 3.3 related to the prediction performance of the applied prediction models, which cannot be captured by the theoretical functions.

Table 4 shows the probabilistic verification measure CRPSS together with the deterministic goodness-of-fit measure KGE′ and its components. These are indicators of the accuracy of the expected value of the predictive distributions. Noting that the CRPS is the integral of the BS over all threshold values, the CRPSS values in Table 4 confirm the results of the BSS analysis above. At both gauges, the skill scores of the different methods are similar. A comparison of the results with goodness-of-fit measures for model predictions in Table 2 reveals that all methods lead to a significant improvement of model predictions in deterministic terms. By focusing on the bias ratio, it becomes evident that the post-processing methods also act as bias-correction. At gauge Trier, also the variability of the expected value of the predictive distributions is very similar to that of observations. In general, all methods perform similarly based on the verification measures in Table 4 with no particular method clearly outperforming the other.

Table 4. Deterministic Goodness-of-Fit (Gof) measure KGE′ and its components of the expected values of the predictive distributions estimated with the COpula uncertainty Processor (COP), Bayesian Model Averaging (BMA), Quantile Regression (QR), Multivariate Truncated Normal (MTN) distribution and the probabilistic skill score CRPSS of the predictive distributions for the gauges Bollendorf and Trier.

GoF Measure	Bollendorf				Trier			
	COP	**BMA**	**QR**	**MTN**	**COP**	**BMA**	**QR**	**MTN**
KGF′	0.91	0.90	0.90	0.93	0.96	0.96	0.96	0.96
Correlation ρ	0.94	0.94	0.94	0.94	0.97	0.97	0.97	0.97
Bias-Ratio β	1.03	1.05	0.99	1.00	1.00	1.01	1.00	1.00
CV-Ratio γ	0.95	0.94	0.92	0.95	0.99	0.98	0.97	0.98
CRPSS	0.61	0.62	0.63	0.63	0.71	0.71	0.72	0.72

To highlight the effect of different model ensemble combinations on PU estimation, we show the verification results of the predictive distribution at gauge Trier using COP for different model combinations in Table 5. It is clear that the combination HBV-ESN and ESN-REW performs similarly compared to the case in which all three models are used. Adding another model prediction improves the results only marginally. When comparing the results of only HBV against the HBV-REW combination, there is no significant increase in performance, implying that the REW model prediction does not add information with respect to HBV. On the other hand, the performance increases considerably when adding the ESN model to either HBV or REW, indicating that the completely different type of model structure, which is of statistical black-box type, contains complementary information.

Table 5. Deterministic Goodness-of-Fit (GoF) measure KGE′ and its components of the expected values of the predictive distributions estimated with the copula uncertainty processor with different combinations of model predictions and the probabilistic skill score CRPSS of the predictive distributions for the gauge Trier.

GoF Measure	HBV REW ESN	HBV	ESN	REW	HBV ESN	HBV REW	ESN REW
KGE′	0.96	0.95	0.95	0.86	0.96	0.95	0.96
Correlation ρ	0.97	0.95	0.96	0.90	0.96	0.96	0.96
Bias-Ratio β	1.00	1.00	1.00	1.01	1.00	1.00	1.00
CV-Ratio γ	0.99	0.98	0.97	0.90	0.99	0.98	0.97
CRPSS	0.71	0.64	0.66	0.49	0.70	0.66	0.69

Figure 8 illustrates the uncertainty bands of the different methods applied for the winter 2002/2003 flood at gauge Trier. The REW model overestimates the flood peak, while ESN underestimates it. In this case, the HBV prediction represents the flood event almost perfectly. Besides the large differences between individual model predictions, the observed values fall well within the 5%–95% uncertainty range for all PU methods applied. The uncertainty bandwidths show the PU densities of QR and COP are sharper with respect to those of the other two methods. The wider uncertainty bands of BMA and MTN confirm the results of the sharpness verification. In summary, we conclude that all methods provide a satisfactory PU estimate for this event.
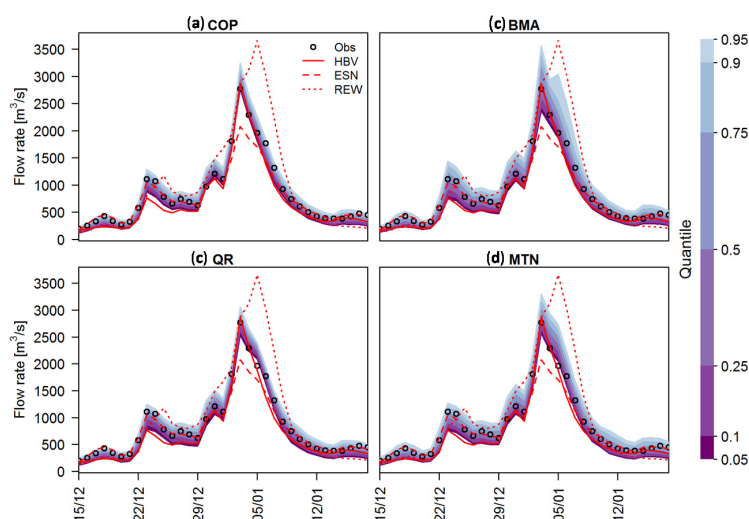

Figure 8. Predictive uncertainty of the flood event 15 December 2002–18 January 2003 at the gauge Trier estimated with the COpula uncertainty Processor (COP) (**a**), Bayesian Model Averaging (BMA) (**b**), quantile regression (QR) (**c**) and Multivariate Truncated Normal (MTN) distribution (**d**). Model predictions of HBV (continuous), ESN (dashed) and REW (dotted) as red lines.

4. Conclusions

The assessment of predictive uncertainty (PU) has become increasingly important in hydrological forecasting over recent years. For this purpose, a variety of methods have been developed and applied to estimate the PU of variables of interest, conditional on hydrological model predictions. In this paper, we introduced a COpula uncertainty Processor (COP), which uses the concept of pair-copula construction to model the multivariate distribution of observation and model predictions. It allows to adequately model different dependence structures among random variables. For the required estimation of the marginal distributions of the variables, we apply mixture probability distributions, given that the flow rates depend on particular hydro-meteorological situations.

We apply the COP to the estimation of the PU for a hydrological and statistical multi-model ensemble at the gauges Bollendorf and Trier in the Moselle river basin. Predictions by the models HBV and REW with conceptual and process-based representation, respectively, as well as the artificial recurrent neural network model Echo State Networks (ESN) serve as random predictors for conditioning PU estimation.

The verification of the quality of the predictive distribution estimated with the COP and the comparison with the results of three other uncertainty processors—BMA (1); QR (2); and MTN (3)—reveals that the COP is very well suitable to estimate the PU. Especially in terms of sharpness, the performances of the methods show differences. The predictive distributions of the COP and the QR are considerably sharper than BMA and MTN. The wider uncertainty bands of the BMA can be explained by use of constant variance for the predictive distributions of individual models, which cannot account for heteroscedasticity of the error variance, and the superposition of the predictive distributions. The larger the differences between the individual model predictions the wider the uncertainty distribution. Results of the sharpness verification are confirmed in the reliability verification via the "Probability Integral Transform" (PIT)-Histogram. Both PIT-Histograms of MTN and BMA indicate overdispersion, which means that the estimated variance of the PU distribution is too large. Due to the calibration of a linear regression model for each quantile, the QR approach leads to best results in terms of reliability verification. This advantage is simultaneously the most significant drawback of the approach because of the large number of fitting parameters required and therefore potentially lower robustness. In summary, the COP provides sharp and yet reliable predictive distributions and considerably more parsimonious than the similarly performing QR. Depending on the selected bivariate copula families, 12 parameters need to be estimated at most for the PU assessment, conditional on three model predictions. Due to the construction of the multivariate probability by independent selected pair-copulas, COP is a flexible and robust approach to estimate the PU of the observation conditional on several model predictions with heterogeneous dependence structures.

Comparison of the effects of different combinations of the hydrological and statistical multi-model ensemble on PU estimation revealed that the results are significantly improved when the purely statistical model was added to the mix of hydrological models, whereas the two deterministic hydrological models hold similar information about the PU. Hence, different kinds of models should be considered in a multi-model approach with the aim to maximize the information content for PU assessment.

In future work, the presented pair-copula approach will be applied to estimate the predictive distribution of deterministic and hydro-meteorological ensemble forecasts for individual forecast values as well as to the multi-temporal case to analyze temporal dependencies between consecutive forecast values.

As the error structure of hydrological model predictions and forecasts is strongly event-dependent—*i.e.*, the model error structure in modeling snowmelt is very different from modelling heavy rainfall events—we aim at using different copula models estimated for particular hydrological events. In this context, the PU estimation will also be made dependent on the specific event type to be analyzed.

Author Contributions: B.K. and D.M. conceived and conducted the study, performed the numerical modelling and wrote the paper. H.-U.K. implemented the ESN model for the two gauges in the Moselle river basin and provided useful suggestions to the manuscript. P.R. implemented the REW model for the Moselle river basin, provided useful suggestions and detailed revision of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Krzysztofowicz, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **1999**, *35*, 2739–2750. [CrossRef]
2. Todini, E. A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manag.* **2008**, *6*, 123–137. [CrossRef]
3. Todini, E. History and perspectives of hydrological catchment modelling. *Hydrol. Res.* **2011**, *42*, 73–85. [CrossRef]
4. Todini, E. The role of predictive uncertainty in the operational management of reservoirs. In Proceedings of the ICWRS2014—Evolving Water Resources Systems: Understanding, Predicting and Managing Water-Society Interactions, Bologna, Italy, 4–6 June 2014.
5. Ramos, M.H.; van Andel, S.J.; Pappenberger, F. Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2219–2232. [CrossRef]
6. Verkade, J.S.; Werner, M.G.F. Estimating the benefits of single value and probability forecasting for flood warning. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3751–3765. [CrossRef]
7. Krzysztofowicz, R. Bayesian system for probabilistic river stage forecasting. *J. Hydrol.* **2002**, *268*, 16–40. [CrossRef]
8. Krzysztofowicz, R.; Kelly, K.S. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* **2000**, *36*, 3265–3277. [CrossRef]
9. Krzysztofwicz, R.; Maranzano, C.J. Bayesian system for probabilistic stage transition forecasting. *J. Hydrol.* **2004**, *299*, 15–44. [CrossRef]
10. Reggiani, P.; Renner, M.; Weerts, A.H.; van Gelder, P. Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system. *Water Resour. Res.* **2009**, *45*, 1–14. [CrossRef]
11. Coccia, G.; Todini, E. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3253–3274. [CrossRef]
12. Todini, E.; Coccia, G.; Ortiz, E. On the proper use of ensembles for predictive uncertainty assessment. Geophysical Research Abstracts 2015. In Proceedings of the EGU2015, Vienna, Austria, 12–17 April 2015.
13. Weerts, A.H.; Winsemius, H.C.; Verkade, J.S. Estimation of predictive hydrological uncertainty using quantile regression: Examples from the national flood forecasting system (England and Wales). *Hydrol. Earth Syst. Sci.* **2011**, *15*, 255–265. [CrossRef]
14. Lopez, P.L.; Verkade, J.S.; Weerts, A.H.; Solomatine, D.P. Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn river: A comparison. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 3411–3428. [CrossRef]
15. Raftery, A.E.; Gneiting, T.; Balabdaoui, F.; Polakowski, M. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **2005**, *133*, 1155–1174. [CrossRef]
16. Duan, Q.Y.; Ajami, N.K.; Gao, X.G.; Sorooshian, S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* **2007**, *30*, 1371–1386. [CrossRef]
17. Vrugt, J.A.; Robinson, B.A. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.* **2007**, *43*, 1–15. [CrossRef]
18. Madadgar, S.; Moradkhani, H. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resour. Res.* **2014**, *50*, 9586–9603. [CrossRef]
19. Gneiting, T.; Raftery, A.E.; Westveld, A.H.; Goldman, T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Mon. Weather Rev.* **2005**, *133*, 1098–1118. [CrossRef]

20. Hemri, S.; Lisniak, D.; Klein, B. Multivariate post-processing techniques for probabilistic hydrological forecasting. *Water Resour. Res.* **2015**. [CrossRef]

21. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Academic Press: Oxford, UK, 2011.

22. Madadgar, S.; Moradkhani, H.; Garen, D. Towards improved post-processing of hydrologic forecast ensembles. *Hydrol. Process.* **2014**, *28*, 104–122. [CrossRef]

23. Brown, J.D.; Seo, D.J. Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. *Hydrol. Process.* **2013**, *27*, 83–105. [CrossRef]

24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.

25. Bergström, S. The hbv model. In *Computer Models of Watershed Hydrology*; Singh, V.P., Ed.; Water Resources Publications: Highlands Ranch, CO, USA, 1995; pp. 443–476.

26. Reggiani, P.; Rientjes, T.H.M. Flux parameterization in the representative elementary watershed approach: Application to a natural basin. *Water Resour. Res.* **2005**, *41*. [CrossRef]

27. Reggiani, P.; Rientjes, T.H.M. Closing horizontal groundwater fluxes with pipe network analysis: An application of the REW approach to an aquifer. *Environ. Model. Softw.* **2010**, *25*, 1702–1712. [CrossRef]

28. Jaeger, H.; Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **2004**, *304*, 78–80. [CrossRef] [PubMed]

29. Lukoševičius, M.; Jaeger, H. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **2009**, *3*, 127–149. [CrossRef]

30. Favre, A.C.; El Adlouni, S.; Perreault, L.; Thiemonge, N.; Bobee, B. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* **2004**, *40*. [CrossRef]

31. Bardossy, A.; Pegram, G.G.S. Copula based multisite model for daily precipitation simulation. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 2299–2314. [CrossRef]

32. Klein, B.; Pahlow, M.; Hundecha, Y.; Schumann, A. Probability analysis of hydrological loads for the design of flood control systems using copulas. *J. Hydrol. Eng.* **2010**, *15*, 360–369. [CrossRef]

33. Kao, S.C.; Govindaraju, R.S. A copula-based joint deficit index for droughts. *J. Hydrol.* **2010**, *380*, 121–134. [CrossRef]

34. Joe, H. *Multivariate Models and Dependence Concepts*; Chapman and Hall: New York, NY, USA, 1997.

35. Nelsen, R.B. *An Introduction to Copulas*; Springer: New York, NY, USA, 1999.

36. Salvadori, G.; De Michele, C.; Kottegoda, N.T.; Rosso, R. *Extremes in Nature: An Approach Using Copulas*; Springer: Dordrecht, The Netherlands, 2007; Volume 56.

37. Sklar, A. *Fonctions de Répartition à n Dimensions et Leurs marges*; Publications de l'Institut de Statistique de L'Université de Paris: Paris, France, 1959; Volume 8, pp. 229–231.

38. Joe, H. Families of m-variate distributions with given margins and m(m − 1)/2 bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics*; Ruschendorf, L., Schweizer, B., Taylor, M.D., Eds.; Institute of Mathematical Statistics: Hayward, CA, USA, 1996; Volume 28, pp. 120–141.

39. Bedford, T.; Cooke, R.M. Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intell.* **2001**, *32*, 245–268. [CrossRef]

40. Bedford, T.; Cooke, R.M. Vines—A new graphical model for dependent random variables. *Ann. Stat.* **2002**, *30*, 1031–1068. [CrossRef]

41. Kurowicka, D.; Cooke, R.M. *Uncertainty Analysis with High Dimensional Dependence Modelling*; John Wiley & Sons: Chichester, UK, 2006.

42. Graeler, B.; van den Berg, M.J.; Vandenberghe, S.; Petroselli, A.; Grimaldi, S.; De Baets, B.; Verhoest, N.E.C. Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1281–1296. [CrossRef]

43. Xiong, L.H.; Yu, K.X.; Gottschalk, L. Estimation of the distribution of annual runoff from climatic variables using copulas. *Water Resour. Res.* **2014**, *50*, 7134–7152. [CrossRef]

44. Vernieuwe, H.; Vandenberghe, S.; De Baets, B.; Verhoest, N.E.C. A continuous rainfall model based on vine copulas. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 2685–2699. [CrossRef]

45. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**, *44*, 182–198. [CrossRef]

46. Brechmann, E.C.; Schepsmeier, U. Modeling dependence with C- and D-vine copulas: The R-package CDvine. *J. Stat. Softw.* **2013**, *52*, 1–27. [CrossRef]

47. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]

48. Cherubini, U.; Luciano, E.; Vecchiato, W. *Copula Methods in Finance*; Wiley: Chichester, UK, 2004.

49. Genest, C.; Rivest, L.-P. Statistical inference procedures for bivariate archimedean copulas. *J. Am. Stat. Assoc.* **1993**, *88*, 1034–1043. [CrossRef]

50. Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B* **1964**, *26*, 211–243.

51. Wang, Q.J.; Shrestha, D.L.; Robertson, D.E.; Pokhrel, P. A log-sinh transformation for data normalization and variance stabilization. *Water Resour. Res.* **2012**, *48*, 1–7. [CrossRef]

52. Van der Waerden, B.L. Order tests for two-sample problem and their power I. *Indag. Math.* **1952**, *14*, 453–458. [CrossRef]

53. Van der Waerden, B.L. Order tests for two-sample problem and their power II. *Indag. Math.* **1953**, *15*, 303–310. [CrossRef]

54. Van der Waerden, B.L. Order tests for two-sample problem and their power III. *Indag. Math.* **1953**, *15*, 311–316. [CrossRef]

55. Bogner, K.; Pappenberger, F.; Cloke, H.L. Technical note: The normal quantile transformation and its application in a flood forecasting system. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 1085–1094. [CrossRef]

56. Hundecha, Y.; Pahlow, M.; Schumann, A. Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. *Water Resour. Res.* **2009**, *45*, 1–15. [CrossRef]

57. Furrer, E.M.; Katz, R.W. Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resour. Res.* **2008**, *44*, 1–13. [CrossRef]

58. Solari, S.; Losada, M.A. A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resour. Res.* **2012**, *48*, 1–15. [CrossRef]

59. Rao, A.R.; Hamed, K.H. *Flood Frequency Analysis*; CRC Press: Boca Raton, FL, USA, 2010.

60. Koenker, R. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.

61. Koenker, R. Quantile Regression in R: A Vignette. Available online: http://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf (accessed on 21 December 2015).

62. Fraley, C.; Raftery, A.; Gneiting, T.; Sloughter, M.; Berrocal, V. Probabilistic weather forecasting in R. *R J.* **2011**, *3*, 55–63.

63. Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2007**, *69*, 243–268. [CrossRef]

64. Kling, H.; Fuchs, M.; Paulin, M. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* **2012**, *424*, 264–277. [CrossRef]