MDPI

Article

Towards an Extension of the Model Conditional Processor: Predictive Uncertainty Quantification of Monthly Streamflow via Gaussian Mixture Models and Clusters

Jonathan Romero-Cuellar [1,2,*], Cristhian J. Gastulo-Tapia [1], Mario R. Hernández-López [1], Cristina Prieto Sierra [3] and Félix Francés [1]

[1] Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; cgastulo@unibagua.edu.pe (C.J.G.-T.); maherlo@alumni.upv.es (M.R.H.-L.); ffrances@upv.es (F.F.)

[2] Agricultural Engineering Department, Universidad Surcolombiana, Avenida Pastrana—Cra 1, Huila 410001, Colombia

[3] IHCantabria—Instituto de Hidráulica Ambiental, Universidad de Cantabria, 39011 Santander, Spain; prietoc@unican.es

[*] Correspondence: jonathan.romero@usco.edu.co

Abstract: This research develops an extension of the Model Conditional Processor (MCP), which merges clusters with Gaussian mixture models to offer an alternative solution to manage heteroscedastic errors. The new method is called the Gaussian mixture clustering post-processor (GMCP). The results of the proposed post-processor were compared to the traditional MCP and MCP using a truncated Normal distribution (MCPt) by applying multiple deterministic and probabilistic verification indices. This research also assesses the GMCP's capacity to estimate the predictive uncertainty of the monthly streamflow under different climate conditions in the "Second Workshop on Model Parameter Estimation Experiment" (MOPEX) catchments distributed in the SE part of the USA. The results indicate that all three post-processors showed promising results. However, the GMCP post-processor has shown significant potential in generating more reliable, sharp, and accurate monthly streamflow predictions than the MCP and MCPt methods, especially in dry catchments. Moreover, the MCP and MCPt provided similar performances for monthly streamflow and better performances in wet catchments than in dry catchments. The GMCP constitutes a promising solution to handle heteroscedastic errors in monthly streamflow, therefore moving towards a more realistic monthly hydrological prediction to support effective decision-making in planning and managing water resources.

Keywords: uncertainty analysis; water resources; cluster analysis; Gaussian mixture model; probabilistic prediction

1. Introduction

Hydrological predictions are beneficial for water management and planning, such as arranging hydraulic infrastructure (irrigation and draining systems, aqueducts, reservoirs, among others), managing flood and drought risks, and estimating ecological flows, operations, and monitoring existing systems—among others [1]. In addition, estimating the predictive uncertainty of monthly streamflow plays a crucial role in supporting decision-making for water resources management, such as water supply, hydropower, and water balance [2]. Moreover, decision-making in the context of water resources is a complex practice due to the investments, the large scale, and the meaning of projects [3]. Furthermore, such hydrologic predictions are affected by various sources of uncertainty, mainly in observed data [4], the model's parameters [5,6], the model's structure [7,8], the initial conditions [9], the model's numerical solution [10], and the intrinsic non-deterministic performance of the systems [11]. Accordingly, Predictive Uncertainty Quantification (PUQ)

is a fundamental tool for risk management and for supporting decision-making in an informed manner when administering water resources [12].

Predictive uncertainty is the probability of observations conditioned by all information and available knowledge (predictions) occurring until today [13]. Therefore, predictive uncertainty is conditioned on the model's structure, parameters, and input data [14]. Hence, PUQ is crucial for making reliable, sharp, and accurate hydrological predictions. It also characterises all the possible predictions and their respective occurrence probabilities [15]. This way of characterising uncertainty does not simplify the decision-making process but provides valuable information about what is not known in the system [16,17]. According to Prieto et al. [18], making predictions without quantifying uncertainty is not knowing reality. Hydrological post-processing methods are suitable for estimating the predictive uncertainty of deterministic hydrological predictions (point predictions).

Formally speaking, a hydrological post-processor is a statistical model employed to improve deterministic predictions by relating the hydrological model's outputs with the corresponding observations [19]. In practice, post-processors are used to characterise the hydrological model's uncertainty and to eliminate the systematic bias of predictions [20]. Post-processors are in charge of mitigating errors in the model's input and output data, parameters, initial conditions, boundary conditions, and structure. Hydrological post-processors have two main objectives: (i) estimating the predictive uncertainty of the hydrological model's deterministic outputs. In this context, hydrological post-processing can be understood as a simple method to convert deterministic predictions into probabilistic ones [21,22]; (ii) correcting the systematic bias of hydrological models to make more accurate predictions.

In recent years, different methods have been developed to estimate the predictive uncertainty of hydrological forecasts. To determine the structure of dependence between the predictions and observations, most methods are based on the meta-Gaussian model, owing to the statistical goodness and facilities that Gaussian variables present [13,14,23–26]. This procedure distributes bivariate probability distributions between deterministic predictions and observations. The errors of hydrological predictions are generally non-Gaussian, heteroscedastic, and autocorrelated [27–30]. To solve this problem, many post-processors apply normalisation methods, such as Normal Quantile Transform (NQT) [31], Box-Cox transformation [32], log-sinh transformation [33], etc.

The first work about predictive uncertainty and hydrological post-processing was conducted by Krzysztofowicz [13] in the context known as the Bayesian Forecasting Framework (BFS). This method developed a bivariate meta-Gaussian distribution function based on a Normal quantile transformation of two variables: observations and predictions according to Gaussian laws. This procedure is known as the Hydrological Uncertainty Processor (HUP) [34]. One of the disadvantages of the HUP is that it does not suitably represent the heteroscedasticity of the error variance. Todini [14] proposed the Model Conditional Processor (MCP), which employs a meta-Gaussian model to estimate the predictive uncertainty of one or a combination of many hydrological models. Coccia and Todini [35] extended the MCP by using Multivariate Truncated Normal distributions to model the joint distribution for many variables in the Normal space to solve the heteroscedastic error problem. Weerts et al. [36] applied quantile regression (QR) to deal with the heteroscedasticity of the hydrological variables' error. QR offers the advantage of analysing the relationship between the observations and predictions from different quantiles, which could be very important for understanding extreme data and managing data with heteroscedasticity [37]. Nonetheless, QR separately estimates one regression for each quantile, generating many parameters.

Similarly, Raftery et al. [38] introduced the Bayesian Model Average (BMA) method that uses many models. Uncertainty is estimated as the average weight of each model's predictive distribution [39]. BMA offers the disadvantage of uncertainty, being conditioned to the number of employed models and their diversification to represent the state variable's uncertainty.

Other hydrological post-processing methods have been implemented, and most employ Bayesian principles. For example, Wang et al. [25] presented the Bayesian Joint Probability (BJP) and Zhao et al. [26] introduced the General Linear Model Post-processor (GLMPP). There are also hydrological post-processors with different error models that had been evaluated under different climate conditions [40,41], such as post-processors that employ non-parametric methods [42], post-processors based on machine-learning principles [43–48], and post-processors based on the copula concept to establish the relation of the dependence among state variables [49–51]. This list of hydrological post-processors is not long, and readers can find more details in the work by Li et al. [52]. Likewise, for reviews of advances in uncertainty analysis, see Moges et al. [53] and Matott et al. [54].

The Model Conditional Processor (MCP) has been established as a hydrologic post-processor for quantifying predictive uncertainty in diverse applications. For instance, precipitation and temperature re-analyses [55], floods in real-time [56], ensemble predictions [57], and satellite rainfall information [58]. Although Coccia and Todini [35] provide insights to deal with the heteroscedastic error using multivariate truncated Normal distributions, the problem is still an open question, especially in monthly streamflow. This paper introduces the Gaussian mixture model and cluster method as a promising alternative to deal with the heteroscedasticity problem, namely that the forecast uncertainty increases with the magnitude of forecast variables.

Nowadays, the use of clusters has become popular in hydrology. For example, Parviz and Rasouli [59] made rain forecasts by artificial intelligence and cluster analysis; Yu et al. [60] implemented the regionalisation of hydroclimate variables with clustering; Basu et al. [61] worked with clusters to analyse floods; and Zhang et al. [62] used clusters and climate similarities to calibrate hydrological models, among others. Likewise, some studies use the Gaussian mixture to represent errors of hydrological variables. For example, Schaefli et al. [63] used a mixture of Normal distributions for quantifying hydrological modelling errors, Smith et al. [64] proposed a mixture of the likelihood for improved Bayesian inference of ephemeral catchments, and Li et al. [65] developed the Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting, which used a sequence of simple error models through four stages. Other authors employ the Gaussian mixture to estimate marginal probability distributions. Thus, Klein et al. [51] proposed a hydrological post-processor based on the bivariate Pair-copula concept and recommended the Gaussian mixture to estimate marginal distributions. Feng et al. [66] introduced a minor modification into the traditional HUP using the Gaussian mixture to estimate marginal distributions. Also, Yang et al. [67] proposed a Bayesian ensemble forecast method, comprising of a Gaussian mixture model (GMM), a hydrological uncertainty processer (HUP), and an autoregressive (AR) model. Finally, Kim et al. [68] used Gaussian mixture clustering to determine groundwater pollution by anthropic effects. It is important to notice that many Gaussian mixture applications were used to estimate marginal distributions.

The importance of estimating uncertainty and support for decision-making in water resources management and planning is stressed. When managing water resources, the monthly temporal discretisation scale is essential for planning the rules for operating in reservoirs, estimating the water balances of catchments, and administrating hydraulic infrastructure in the long term. To deal with these problems, monthly streamflow was employed to evaluate hydrological post-processing.

This paper develops an extension of the MCP [14], which merges clustering with the Gaussian mixture model to offer an alternative solution to manage heteroscedastic errors. The new method is called the Gaussian mixture clustering post-processor (GMCP). The results of the proposed post-processor were compared to the MCP [14] and the MCPt [35] by applying multiple deterministic and probabilistic verification indices. This research also assesses the GMCP's capacity to estimate the predictive uncertainty of the monthly streamflow under different climate conditions in the 12 MOPEX catchments [69] distributed in the SE part of the USA.

To achieve the above goals, this paper is structured as follows: the reference hydrological post-processing methods and the basis for the new post-processor are described; next, the origin of the hydrological predictions and the characteristics of the 12 MOPEX Project catchments are presented to prove the new post-processor's predictive performance; this is followed by the Results, Discussion, and Conclusions sections. All the analyses were carried out using the R Statistical software [70].

2. Materials and Methods

The GMCP post-processor is a statistical model to transform point predictions obtained by any deterministic model into probabilistic predictions, thus deriving the predictive uncertainty of the predictand. The GMCP computes the probability distribution of the observed data conditioned on the generic deterministic model's output (point predictions), along with its mode (median or mean) value and uncertainty band, which is asymptotically consistent in quantifying total uncertainty. In general, all MCP post-processors assessed are based on the following main assumptions:

1. Uncertainty of weather forecasts has been substantially reduced because past observations are employed as the hydrological model's input.
2. Predictions and observations correlating, and this system performance will continue in the future. Similarly, modelled variables are stationary during the calibration and application period. Non-stationarity can be accounted for using deterministic model non-stationarity [71,72]. Such extension is not considered in the present contribution, but a discussion is provided in Section 4.
3. A single deterministic model with a single parameter set is considered. Section 4 will discuss the possible extension of the GMCP post-processor to multi-model applications.
4. The calibration dataset is long enough to ensure sufficient information to upgrade the deterministic and post-processor models. The predictive capacity of the models is limited by proper calibration, which implies that sufficiently long records of observed data, guiding to a variety of hydrologic conditions, are available for model training.

As previously mentioned, this research aims to develop an extension of the MCP [14], which merges clusters with a Gaussian mixture model to offer an alternative solution to manage heteroscedastic errors. The method is identified with the acronym "GMCP" post-processor. This research also assesses the GMCP's capacity to estimate the predictive uncertainty of the monthly streamflow under different climate conditions in 12 catchments in the MOPEX Project [69]. The results of the proposed post-processor were compared to the MCP [14] and the MCPt [35].

2.1. Predictive Uncertainty

In hindcasting, predictive uncertainty describes the probability of any value conditioned to all the information and knowledge acquired by hydrological predictions [13,14,16]. Krzysztofowicz [13] and Todini [14] emphasise two basic theses. Firstly, the objective of hydrological predictions is to quantify the uncertainty of observations rather than the uncertainty of hydrological models. Secondly, the main aim to improve hydrological predictions is to estimate the actual streamflow and to reduce their predictive uncertainty. To better explain these ideas, and to keep in line with Todini [14], a joint probability distribution concept of observations q_o and predictions q_s is presented. Figure 1 shows the joint sample's frequency of q_o and q_s that can be used to quantify the joint probability density function. For any given hydrological model, predictions q_s should be a function of the model parameters (θ) and the input data (x) (precipitation, evapotranspiration, and others.) Therefore, joint density probability can be expressed as $f\left(q_o, \left(q_s | x, \hat{\theta}\right)\right)$. To predict q_o, the conditional predictive distribution must be derived from q_o given q_s. This can

be accomplished by conditioning the joint probability density to the q_s predicted value (Figure 2) and renormalizing. This can be formally expressed as:

$$f(q_o|(q_s|x,\hat{\theta})) = \frac{f(q_o,(q_s|x,\hat{\theta}))}{\int f(q_o,(q_s|x,\hat{\theta}))\,dq_o} \tag{1}$$

It is stressed that the conditional predictive uncertainty of Equation (1) represents the predictive uncertainty given a hydrological model, input data, certain conditions, and some hydrological parameters. Accordingly, and for this paper, the term "predictive uncertainty" refers to "conditional predictive uncertainty". As Figure 1 shows, the conditional distribution $f(q_o|q_s)$ is not as dispersed as the marginal distribution is for $f(q_o)$ because uncertainty could be reduced by any further information provided by the hydrological model's predictions.
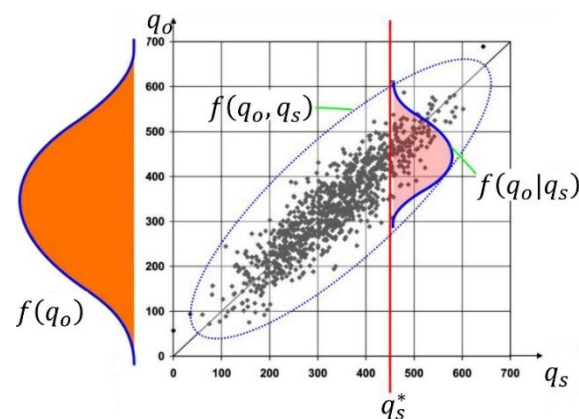
Figure 1. Predictive density is defined as the probability density of the observed variable q_o that is conditional on the hydrological model's predictions, q_s, where q_s is considered to be known in the prediction time (adapted from Todini [14]).
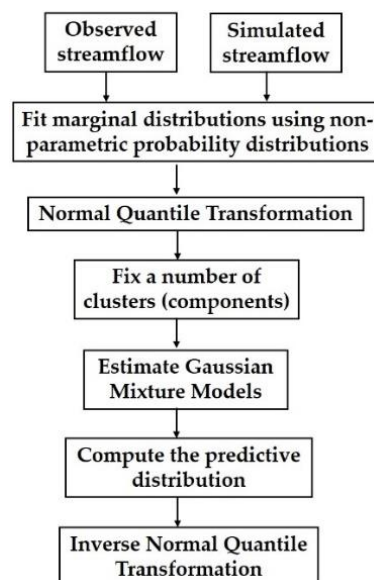
Figure 2. The flow chart of the proposed Gaussian Mixture Clustering Post-processor (GMCP).

2.2. Marginal Distribution and Normal Quantile Transformation

In general, the problem with Gaussian approaches in hydrology is that variables do not tend to be distributed as Gaussian. Therefore, some kind of statistical transformation must be applied to take the hydrological variables to the Gaussian space and to thus adjust

the joint Gaussian probability density distribution (PDF) of both the observations and predictions. The present research applied Normal Quantile Transform (NQT) [31] to all the evaluated post-processing methods. Two auxiliary variables, η_o and η_s, derive from NQT to replace $F(q_o)$ and $F(q_s)$ so that the probability distribution of the observations and predictions in the Gaussian space would, respectively, be:

$$\begin{aligned} \eta_o &= N^{-1}(F(q_o)), \\ \eta_s &= N^{-1}(F(q_s)), \end{aligned} \tag{2}$$

where N represents the standardGaussian distribution with zero mean and unit variance, and $F()$ symbolises marginal distributions. The present research used non-parametric probability distributions to adjust marginal distributions because monthly streamflow is heterogeneous, and the data represent different hydrological situations that might not be easy to describe with the parametric distribution. The kernel density estimation method was applied to adjust the marginal distributions of the random variables [73]. For a random bivariate sample, $X_1, X_2, \ldots, X_n$, are obtained from a joint PDF, f, and the kernel density estimation is defined as:

$$\hat{f}(x; H) = n^{-1} \sum_{i=1}^{n} K_H(x - X_i), \tag{3}$$

where $x = (x_1, x_2)^T$ and $X_i = (X_{i1}, X_{i2})^T$, $i = 1, 2, \ldots, n$. Here, $K(x)$ is the kernel, which is the asymmetric probability density function, H is the symmetric and positive bandwidth matrix, and $K_H(x) = |H|^{-0.5} K(H^{-0.5}x)$. Selecting K is not fundamental: the standard Gaussian distribution $K(x) = (2\pi)^{-1} e^{(-0.5x^T x)}$ was used. Conversely, selecting H is very important for $\hat{f}$ performance [73]. The most widely used parametrization for the bandwidth matrix is the diagonal $H = diag(h_1^2, h_2^2, \ldots, h_n^2)$ with no constraints in H, but it ensures that H is positive and symmetric. For the present research, the kernel estimation was applied using the last square cross-validation method implemented in the ks library [74] of the R statistical software [70].

2.3. Hydrological Post-Processing Methods

The streamflow post-processing methods assessed in this research consist of implementing the Model Conditional Processor (MCP) [14] and some of its ramifications from the MCP using a truncated Normal distribution (MCPt) [35] to finish with the proposed extension of the MCP [14], which merges clustering with a Gaussian mixture model to offer an alternative solution to manage heteroscedastic errors. The new method is called the Gaussian mixture clustering post-processor (GMCP). Only a short overview of the theory of the methods is given here. For future details, we refer to cited publications.

2.3.1. Model Conditional Processor (MCP)

Todini [14] proposed the Model Conditional Processor (MCP), a meta-Gaussian approach initially designed to estimate the predictive uncertainty of floods in real-time. The MCP can be used in several ways: bivariate (observed, simulated), multivariate (several prediction models), unique forecast horizon [35], and multiple lead-time [56]. The MCP is a well-accepted hydrological post-processing method by the hydrological community [55–58]. The MCP mainly establishes a joint probability distribution to describe the relationship between the deterministic hydrological predictions and the corresponding observations. The joint probability distribution is modelled as a bivariate Gaussian distribution, followed by adjusting the marginal distributions and transforming the Gaussian space variables. The MCP, herein employed, includes three steps. The first is the transformation of the predictions and observations to the Gaussian space by the NQT transformation method [31], as shown in Section 2.2. The second step is predictive distribution, which was calculated using Bayes' Theorem by assuming that both the predictions and observations are available simultaneously. In line with the notation of the present paper, observations q_o were

transformed to Gaussian space $\boldsymbol{\eta_o}$ and predictions $\boldsymbol{q_s}$ were transformed into $\boldsymbol{\eta_s}$. Therefore, the relation between $\boldsymbol{\eta_o}$ and $\boldsymbol{\eta_s}$ was formulated using a bivariate Gaussian distribution:

$$\begin{bmatrix} \eta_o \\ \eta_s \end{bmatrix} \sim N(\boldsymbol{\mu},\ \boldsymbol{\Sigma}), \tag{4}$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_{\eta_o} \\ \mu_{\eta_s} \end{bmatrix}$ is the means vector and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{\eta_o}^2 & \rho_{\eta_o \eta_s}\sigma_{\eta_o}\sigma_{\eta_s} \\ \rho_{\eta_o \eta_s}\sigma_{\eta_o}\sigma_{\eta_s} & \sigma_{\eta_s}^2 \end{bmatrix}$ is the covariance matrix. In step three, the predictive uncertainty estimated in Gaussian space was reconverted into real space by the inverse of NQT. The series of observations was divided into two parts to identify the MCP parameters. The first half of the series was used for calibration purposes; that is, to identify marginal distributions and joint distribution thorough Bayes' Theorem. The second half of the series was employed to validate the MCP; the calibrated MCP was conditioned to new predictions to evaluate its performance for a group of parameters $\boldsymbol{\theta}$ and the new predictions were transformed into Gaussian space $\boldsymbol{\eta_{s\_new}}$:

$$\eta_{o\_new}|\eta_{s_{new}},\ \theta \ \sim \ N\left[\mu_{\eta_o} + \rho_{\eta_o \eta_s}\frac{\sigma_{\eta_o}}{\sigma_{\eta_s}}\left(\eta_{s_{new}} - \mu_{\eta_s}\right),\ \sigma_{\eta_o}^2\left(1 - \rho_{\eta_o \eta_s}^2\right)\right], \tag{5}$$

Interestingly, the MCP is simple to implement with a low computational cost because the bivariate Gaussian distribution is analytically processed. Likewise, the parameters are analytically identified, saving the total parametric inference cost. For further details about the MCP, we recommend that readers look at the work of Todini [14]. Next, an improved version of the traditional MCP is presented.

2.3.2. MCP Using Truncated Normal Distribution (MCPt)

To address the heteroscedasticity in the error variance, Coccia and Todini [35] extended the MCP [14] by joining two truncated Normal distributions (TND). The general recommendation is to use two TNDs to characterise the heteroscedasticity of the error variance properly. In line with our monthly streamflow research objective, two TNDs were used; that is, two variances were employed. The split of the Normal multivariate space into two parts is obtained by identifying an M-dimensional hyperplane:

$$H_p = \sum_{i=1}^{M} \eta_{s_i} = M \cdot a, \tag{6}$$

where M is the number of models and η_{s_i} is the prediction in Gaussian space. The threshold a can be distinguished as the value of η_{s_i} that minimizes the predictive variance of the upper sample. In other words, the value of a is identified by minimizing the predictive variance of the upper sample. The predictive uncertainty for the sample above the truncation hyperplane is represented as:

$$f\left(\eta_o \Big| \eta_{s_i} = \eta_{s_i}^*, H_p^* > M*\right) \ \sim \ N\left(\mu_{\eta_o|\eta_{s_i}=\eta_{s_i}^*, H_p^*>M*a},\ \sigma_{\eta_o|\eta_{s_i}=\eta_{s_i}^*, H_p^*>M*a}^2\right) \tag{7}$$

Here, $\eta_{s_i}^*, H_p^*$ symbolize a new realization of predictions and a new hyperplane, respectively. Moreover, the predictive mean is represented by:

$$\mu_{\eta_o|\eta_{s_i}=\eta_{s_i}^*, H_p^*>M\cdot a} = \mu_{\eta_o} + \sum_{\eta_o \eta_s} \sum_{\eta_s \eta_s}^{-1}\left(\eta_{s_i}^* - \mu_{\eta_s}\right), \tag{8}$$

and the predictive variance:

$$\sigma_{\eta_o|\eta_{s_i}=\eta_{s_i}^*, H_p^*>M\cdot a}^2 = \sum_{\eta_o \eta_o} - \sum_{\eta_o \eta_s} \sum_{\eta_s \eta_s}^{-1} \sum_{\eta_o \eta_s}^{T}, \tag{9}$$

In Equations (7)–(9), μ_{η_o} and μ_{η_s} are the sample means, while $\sigma^2_{\eta_o|\eta_{s_i}}$ is the conditional variance and T is the transpose of a matrix. In addition, the model parameters, i.e., the mean, variance, and covariance matrices, are computed from the data of the upper sample. The predictive uncertainty distribution of the lower sample looks similar but is characterised by the values of the sample below the truncation hyperplane $H^*_p \leq M \cdot a$, for more details see [35]. The MCP and MCPt can work with a multi-model, but, for this research, we used only a single model for which we aimed to quantify the total predictive uncertainty. We did not include multiple models because of the ease of understanding and transparency of the procedure. This assumption is further discussed in Section 4.

2.3.3. Gaussian Mixture Clustering Post-Processor (GMCP)

The extension of the MCP method known as GMCP post-processor came about after merging the bivariate Gaussian outline and grouping it into clusters with the Gaussian mixture models (GMM). This means that the GMCP post-processor begins with MCP [14] for a single hydrological model in Section 2.3.1, but the GMCP post-processor offers a different way to deal with the heteroscedasticity of the error variance when the error variance is characterised by clustering with GMMs. The Gaussian mixture is well established in the literature to find homogeneous groups (clusters) in heterogeneous data. The idea of employing GMM to perform a cluster analysis is not new. Wolfe [75] was the first to test GMMs to find clusters. The GMMs offer the advantage of including a probability measure when assigning cluster data. This assignment is known as a soft cluster, where data have a probability of belonging to each cluster [76].

The basic idea behind mixture models of probability distributions to perform cluster analyses consists of assuming that data come from a mixture of underlying probability distributions. The most well-known approach is the Gaussian mixture model (GMM) [77], in which each observation is assumed to be distributed into g Normal distributions and g is the number of clusters (components). For more details, readers refer to the work of Fraley and Raftery [78]. Generally, when GMMs are employed to perform cluster analyses, the same model type is employed ($f_g(x|\theta_g)$) for all the components (clusters), which, in this case, is Gaussian, but with different means and covariance structures.

There are different automatic methods to select the number of mixture components and their parameters [79]. However, the number of mixture components can also be fixed by some prior knowledge about the modelled phenomenon. This research assumes that the joint probability distribution of the observed and simulated data (model error) can be grouped into three categories of variance, and thus choose a three-components Gaussian mixture model. We fixed the number of components a priori to three, thereby corresponding to the high, middle, and low flow period, which is typical of monthly streamflow. Using more than three components is possible, but we will show that three components are sufficient for monthly streamflow and water resources applications in our case studies.

The GMCP provides a semi-parametric outline to model unknown probability distributions, which are represented as the weighted Gaussian sum [80]. Specifically, GMMs possess the flexibility of non-parametric methods with the added advantage of a lower number of parameters, i.e., the dimension of the parameter's vector [81]. To express the mathematical basis of the cluster with GMMs, let us take X as a random vector that stems from the G in the Gaussian mixture distribution. For all $x \in X$, its probability density can be expressed as:

$$f(x|\vartheta) = \sum_{g=1}^{G} \pi_g \, f_g(x|\theta_g) = \sum_{g=1}^{G} \pi_g \, N_g(x|\mu_g, \Sigma_g), \tag{10}$$

where weights $\pi_g > 0$ and $\sum_{g=1}^{G} \pi_g = 1$, which are known as the mixture proportion or weighted weights; $f_g(x|\theta_g)$ is the gth component of probability density; $\theta = \theta_1, \ldots, \theta_G$ and $\pi = \pi_1, \ldots, \pi_G$ are the parameters; N_g represents the Normal distribution; μ_g is the means vector; and Σ_g is the covariance matrix for each component (cluster) g. This research

employed two random variables (observed streamflow (η_o) and simulated streamflow (η_s) after normalisation). Then, each data pair (η_o, η_s) was modelled as if sampled from one of the g probability distributions ($N_1(\mu_1, \Sigma_1), N_2(\mu_2, \Sigma_2), \ldots, N_g(\mu_g, \Sigma_g)$). For example, assuming that three clusters were identified, the probability of belonging to a given cluster lowers as data points (η_o, η_s) move away from the cluster centre.

Now, let us assume that $z_i = (z_{i1}, \ldots, z_{iG})$ represents the membership of the component of observation i. Thus, $z_{ig} = 1$ if observation i belongs to component g, and $z_{ig} = 0$ otherwise. Let us also assume that the n vectors of data $x_1, \ldots, x_n$ are observations with no assigned component g. In this scenario, the likelihood function is:

$$\mathcal{L}(\vartheta|x) = \prod_{i=1}^{n} \sum_{g=1}^{G} \pi_g \, N(x_i|\theta_g), \tag{11}$$

where N represents the Normal probability distribution. The parameters were estimated with the Expectation-Maximization (EM) algorithm [82]. This algorithm is an iterative procedure followed to estimate the maximum likelihood function. Having estimated the parameters, the predictive classification results are supplied by the *a posteriori* probability distribution:

$$\hat{z_{ig}} = \frac{\hat{\pi}_g \, f(x_i|\hat{\theta}_g)}{\sum_{h=1}^{G} \hat{\pi}_h \, f(x_i|\hat{\theta}_h)}, \tag{12}$$

for $i = 1, \cdots, n$. The complete cluster grouping analysis was implemented with GMMs using the *mclust* library [83] of the R statistics software [70]. Figure 2 displays the flow chart of the procedure for applying the GMCP post-processor.

2.4. Case Studies

The data, herein employed, were the observed and simulated monthly streamflow obtained from the "Second Workshop on Model Parameter Estimation Experiment (MOPEX)" [69]. The MOPEX project is a well-known reference database in the international hydrological community that has mainly been used to evaluate hydrological models and theories [8,10,84]. For example, and particular to this paper, Ye et al. [19] used the 12 catchments from the MOPEX database to compare the results from post-processing and calibrated the hydrological models. Thus, MOPEX offers a valuable opportunity to evaluate and compare the performance of new hydrological post-processing methods under different climate conditions. From the MOPEX database, 12 catchments were selected, which are distributed in the SE area of the USA. The Aridity Index (relation between potential evapotranspiration and precipitation) ranges from 0.43 to 2.22, and the Runoff Ratio (relation between surface run-off and precipitation) varies between 0.15 and 0.63 (Table 1). Thus, the 12 catchments selected from the MOPEX project represent different climate conditions (Figure 1). Basic information about them is supplied in Table 1. We selected the same 12 MOPEX catchments used by Ye et al. [19] to discuss the results.

Figure 3 depicts the Budyko curve for all 12 catchments from the MOPEX project. According to the Budyko hypothesis, if the energy available in a catchment suffices to evaporate humidity, then the catchment is limited by water availability (catchment B12 has the highest Aridity Index). Conversely, if the available energy does not suffice to evaporate humidity, the basin is limited by energy availability (catchment B2 is the exact opposite of B12 as it has the lowest Aridity Index). It is worth stressing that the 12 selected catchments were distributed all over the Budyko curve, as Figure 3 depicts, which ensures the critical evaluation of post-processors under different climate conditions.

Table 1. Hydrological information about the 12 catchments selected from the Mopex project.

ID	Station Name	Elev.	Area (km^2)	P	PET	Q	Run-Off Index (Q/P)	Aridity Index (PET/P)
B1	Amite River Near Denham Springs, LA	0	3315	1560	1068.5	612	0.39	0.67
B2	French Broad River at Asheville, NC	594	2448	1378	588.9	795	0.58	0.43
B3	Tygart Valley River at Philippi, WV	390	2372	1164	661.4	736	0.63	0.57
B4	Spring River Near Waco, MO	254	3015	1075	1119.8	300	0.28	1.04
B5	S Branch Potomac River Nr Springfield, WV	171	3810	1043	636	339	0.33	0.61
B6	Monocacy R At Jug Bridge Nr Frederick, MD	71	2116	1042	906.1	421	0.4	0.87
B7	Rappahannock River Nr Fredericksburg, VA	17	4134	1028	856.7	375	0.36	0.83
B8	Bluestone River Nr Pipestem, WV	465	1020	1017	678	419	0.41	0.67
B9	East Fork White River at Columbus, IN	184	4421	1014	838	377	0.37	0.83
B10	English River at Kalona, IA	193	1484	881	989.9	261	0.3	1.12
B11	San Marcos River at Luling, TX	98	2170	819	1462.5	170	0.21	1.79
B12	Guadalupe River Nr Spring Branch, TX	289	3406	761	1691.1	116	0.15	2.22

Elev: elevation (m), P: mean areal precipitation (mm/year), PET: potential evapotranspiration (mm/year), Q: observed streamflow (mm/year).
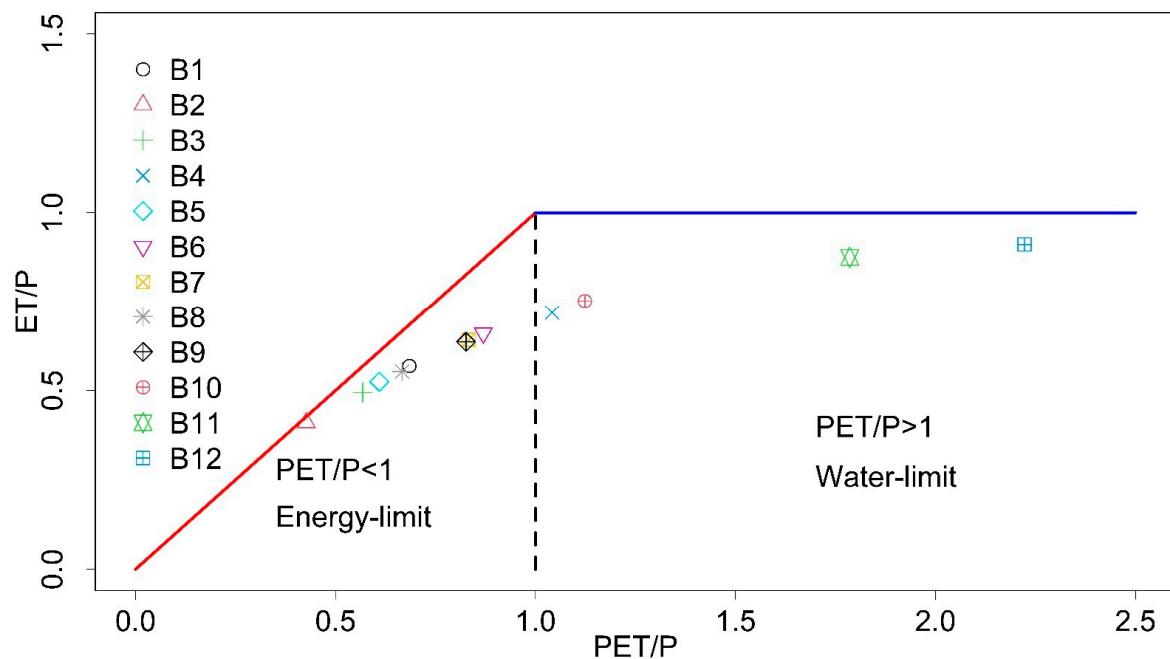
Figure 3. The Budyko curve for the 12 catchments selected from MOPEX. The values to reproduce the figure came directly from the MOPEX database.

2.5. Hydrological Model

The GR4J hydrological model predictions were employed [85], which are a well-known and widely used model in different parts of the world. GR4J is a lumped conceptual model with four calibration parameters: maximum capacity of the production store x_1 (mm); groundwater exchange coefficient x_2 (mm); 1-day-ahead maximum capacity of the routing store x_3 (mm); and time base of unit hydrograph x_4 (days). For further information about the model's description, readers refer to the work by Perrin et al. [85]. Daily predictions were aggregated on a monthly basis to evaluate the post-processors' performance for planning and managing water resources. We want to emphasise that the GR4J hydrological

model predictions were not prioritised because they are input data. According to the aim of this paper, we focused on the performance of the post-processors.

2.6. Verification Indices

Assessment of the performance is vital to offer end-users an indication of the predictions´ reliability and uncertainty bands. Some verification indices exist that can be used to assess the performance of hydrological post-processors. This research employed deterministic and probabilistic verification indices, which evaluate the hydrological predictions´ accuracy, sharpness, and reliability. These indices were also recommended by Laio and Tamea [86], Renard et al. [6], and Thyer et al. [87]. The deterministic Nash–Sutcliffe efficiency index (NSE) [88] was applied to the predictive distribution mean. This index does not assess the complete predictive distribution but is a classic index and a general reference in hydrology. The NSE measures the squared differences between predictions q_s and observations q_o, which are normalised by the variance of the observations:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}(q_s - q_o)^2}{\sum_{i=1}^{n}(q_o - \overline{q_o})^2},$$ (13)

where $\overline{q_o}$ is the average of the observations. Probabilistic indices were employed to assess the predictive distributions. The predictive quantile–quantile (PQQ) plot [86] was applied. This diagram shows how probabilistic predictions represent the observations´ uncertainty [86,87]. If both predictive distribution and observations are consistent in the PQQ context, the value corresponding to the distribution *p*-value must be uniformly distributed throughout the interval $[0,1]$. In other words, predictions are considered reliable when the relative frequency of the observations equals the frequency of predictions. This situation can be visually identified when the PQQ curve follows the bisector (line 1:1). Otherwise, predictive distribution deficiencies can be interpreted when the curve moves away from the bisector. Indeed, according to Laio and Tamea [86], the predictive distribution can display three patterns. If the PQQ plot follows the bisector, the predictive uncertainty is correctly estimated, and the observations are a random sample of the predictive distribution. Conversely, if the PQQ plot shows an "S"-shape, it means that the predictive distribution is underestimated (large bands) and an inverted "S"-shape implies an overestimated uncertainty (narrow bands). From the PQQ plot, we can deduce two indices: reliability and sharpness.

The reliability index quantifies the statistical consistency between the observations and predictive distribution:

$$\text{Reliability} = 1 - \frac{2}{n} \sum_{i=1}^{n} \left| F_U - F_{q_s}(q_o) \right|,$$ (14)

where F_U is a uniform cumulative distribution function (CDF) and $F_{q_s}(q_o)$ is the predictive CDF. The reliability index ranges from 0 (the worst reliability) to 1 (perfect reliability).

The sharpness index is related to the predictive distribution concentration. In other words, it refers to the coverage provided by the distribution [6]:

$$\text{sharpness} = \frac{1}{n} \sum_{i=1}^{n} \frac{\text{E}[q_s]}{\sigma[q_s]},$$ (15)

where $\text{E}[]$ and $\sigma[]$ are the operators of the expected value and standard deviation, respectively. The sharpness index range is $(0, \infty)$, and the predictive distributions with higher sharpness index (narrower) values are more accurate. Predictive distributions can be found with equal reliability indices but different degrees of sharpness, in which case the higher sharpness values are preferable because they denote more accurate predictive distributions.

Furthermore, the containing ratio (95%CR) was used. The 95%CR is the percentage of observations that fall within the 95% uncertainty band. In this research, the 95% band

was estimated to be within the 2.5 and 97.5 percentiles. This allowed the quantification of the desired uncertainty to be achieved when the 95%CR came close to 95%. As the presented verification indices are well-known in the literature, no lengthy description is provided. However, readers are recommended to the works of Franz and Houge [84], Laio and Tamea [86], and Renard et al. [6] for further information.

2.7. Comparison Frame

It should be remembered that the main aim of the present paper is to develop an extension of the MCP [14], which merges clustering with a Gaussian mixture model to offer an alternative solution to manage heteroscedastic errors. In addition, comparing GMCP's performance to similar post-processing methods under different climate conditions is also needed as a benchmark. In order to perform the post-processing of monthly streamflow and to quantify predictive uncertainty, the following procedure was used.

First, daily streamflow predictions were obtained from the GR4J hydrological model [85], and were calibrated and validated by Ye et al. [19] for the 12 MOPEX catchments. Given this, the hydrological model outputs (previously calibrated and validated) become the inputs for the evaluated hydrological post-processors.

Second, the daily hydrological predictions were aggregated monthly because the post-processing methods were applied in the water resources management context.

Third, to evaluate the post-processing methods, the time series of both observations and predictions were divided into 20 years to calibrate the post-processors' parameters (1960–1980) and into 17 years for the validation (1981–1998).

Fourth, NQT [31] was applied to all the evaluated post-processors with non-parametric marginal distributions to map the observations and simulations to the Normal space. The three evaluated post-processors were separately implemented into the 12 MOPEX catchments to find the best performing post-processors. The 12 MOPEX catchments were selected because they were the same catchments employed in previous studies to compare hypotheses, which the hydrological community is very familiar with, e.g., [8,19,84,89].

Moreover, evaluating hydrological post-processors under different climate conditions allows for more general recommendations to be obtained [90].

Finally, evaluating the predictive uncertainty with only one verification index can lead to mistaken interpretations and wrong decision-making for managing water resources [41]. Consequently, many independent verification indices were used together instead of individual ones.

3. Results

The hydrological post-processing methods evaluated according to the framework described in the previous subsection are presented. The results correspond to the validation period, as it is the most critical period where the predictive uncertainty of the analysed methods is identified. Section 3.1 benchmarks the GCMP post-processor with the MCP [14] and MCPt [35] to quantify the predictive uncertainty of the monthly streamflow, which is conditional on deterministic model predictions. The case studies consider 12 MOPEX catchments with a diverse range of hydroclimatology. The Nash–Sutcliffe efficiency index (NSE), sharpness, and the containing ratio (95%CR) verification index are presented in Figure 4. Moreover, the PQQ plots, which assess the reliability, sharpness, and bias, are depicted in Figure 5.

An initial inspection of the results found considerable overlap in the performance verification indices achieved by the MCP and MCPt post-processors for monthly streamflow. The MCP and MCPt also showed poor performance in dry catchments. Conversely, the GMCP post-processor empirically made the most accurate, reliable, and sharpest predictions.
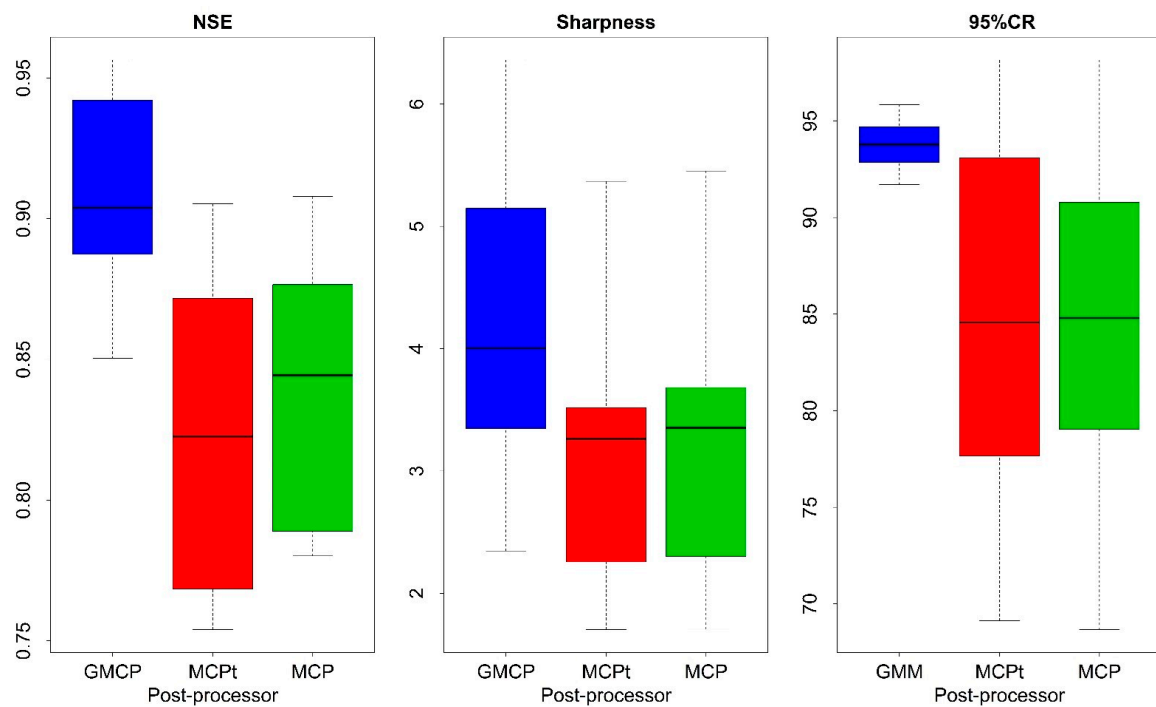
Figure 4. Performance of monthly predictions in terms of NSE, sharpness, and containing ratio (95%CR) for the three post-processors during the validation period (1980–1998) overall catchments.
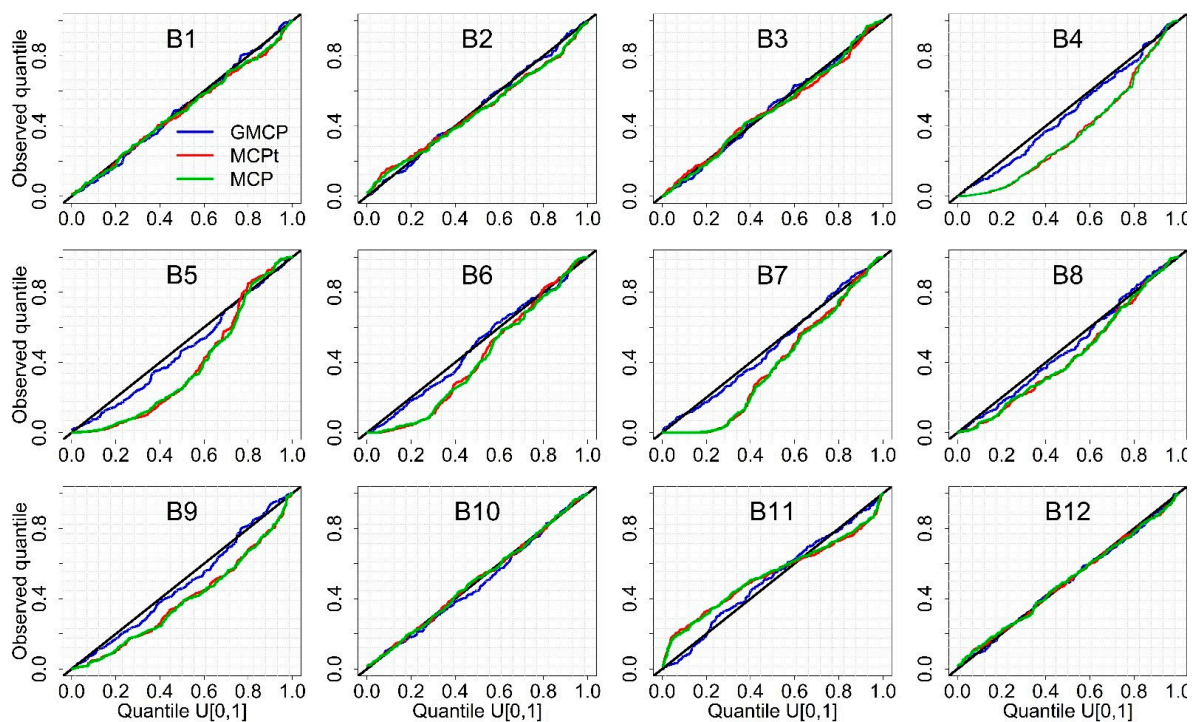

Figure 5. Predictive PQQ plot of the three evaluated post-processors and 12 MOPEX catchments during the validation period (1980–1998).

The streamflow forecast time series and corresponding skill for a single catchment, the San Marcos catchment (B11), are presented in Figure 6. Then, the relation of the Aridity index with the performance of post-processors is shown in Figure 7.
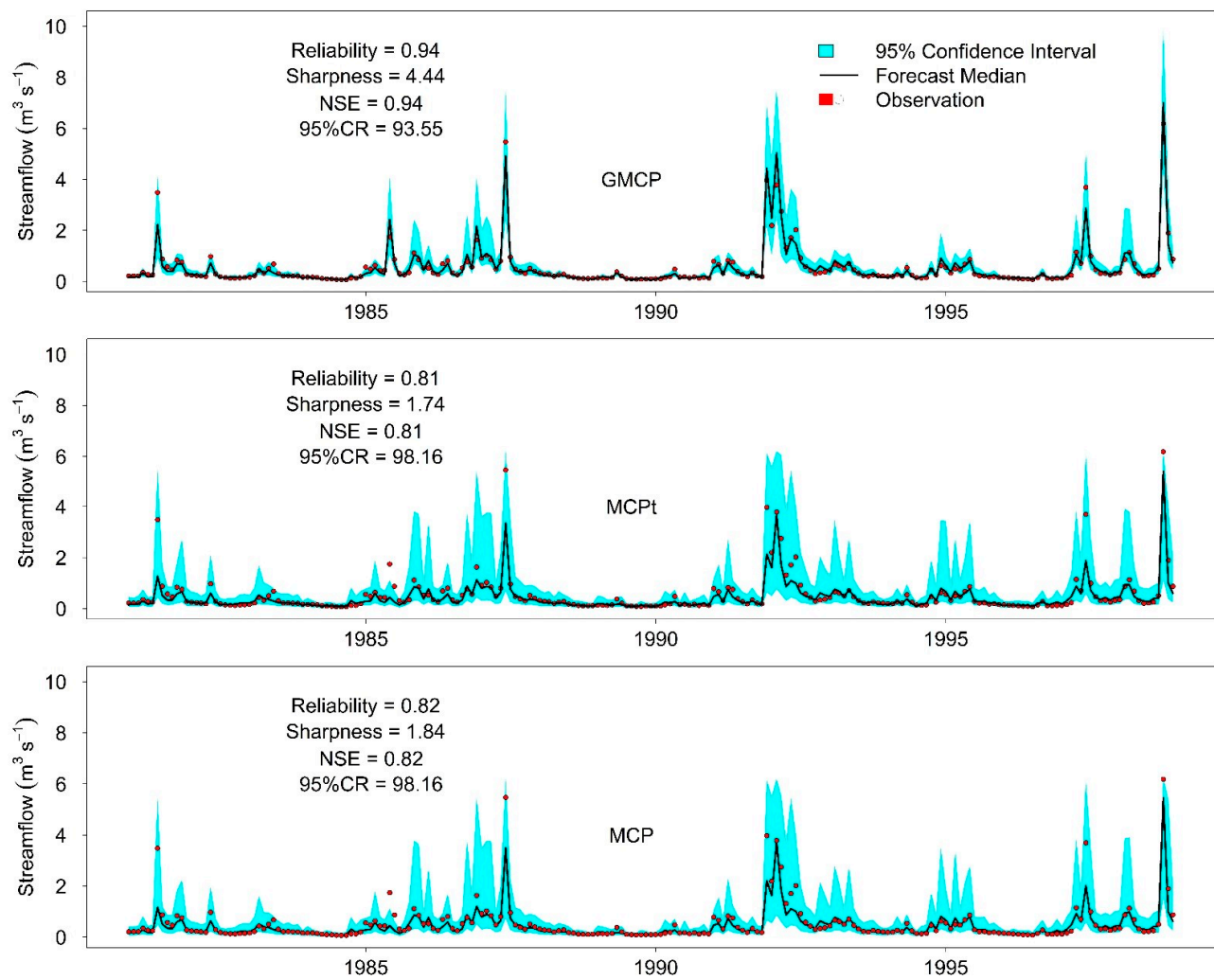
Figure 6. Time series of the median and 95% confidence interval of monthly streamflow predictions derived from the Gaussian mixture clustering post-processor (GMCP), model conditional processor (MCP), and MCP using the truncated Normal (MCPt), compared with observations from the validation period (1980–1998) in the San Marcos catchment (B11).
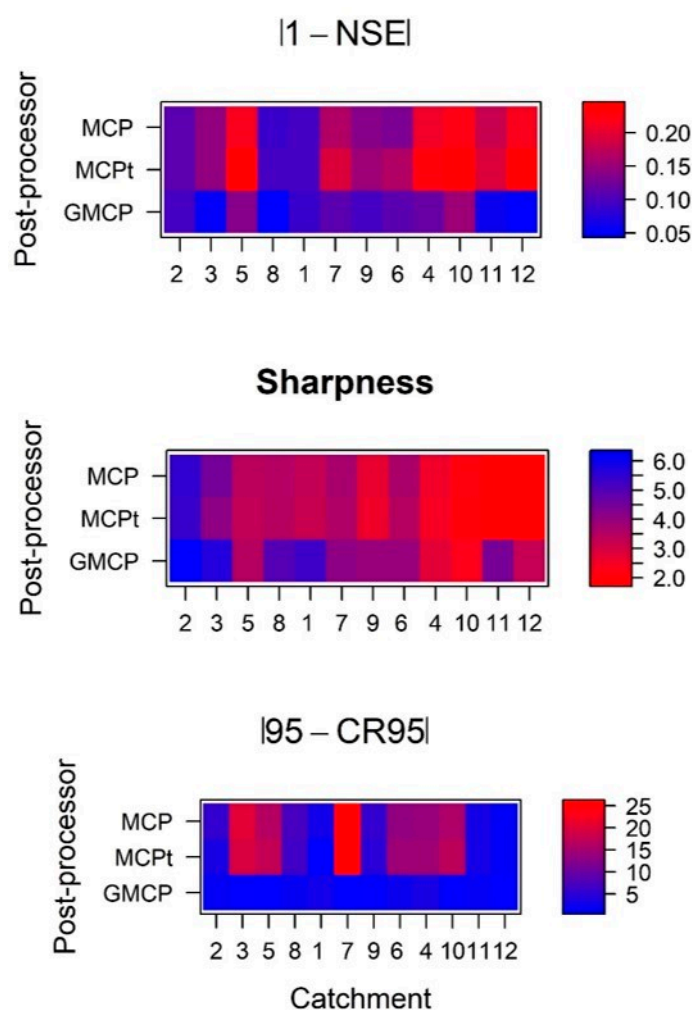
Figure 7. Comparison of the deterministic and probabilistic metrics computed for the three post-processors during the validation period (1980–1998) for 12 MOPEX catchments. MCP (model conditional processor), MCPt (MCP using the truncated Normal) and GMCP (Gaussian mixture clustering post-processor).

3.1. Comparison of Post-Processors: Individual Verification Indices

Figure 4 offers the average values for the verification indices for the 12 catchments in boxplot-type diagrams. In terms of the Nash–Sutcliffe efficiency index (NSE), Figure 4 shows considerable overlap in the boxplots corresponding to MCP and MCPt. This finding suggests little difference in the performance of these post-processors for monthly streamflow. Moreover, the NSE indices are generally suitable for all 12 catchments and assessed post-processors (Figure 4, left panel), and, according to the classification of Martinez and Gupta [91], NSE > 0.75 is considered a good result. Overall, these results suggest the GMCP is consistently better in terms of the NSE values because of its higher NSE indices and shows less dispersion in the boxplot.

Regarding the sharpness index, and in line with Figure 4 (middle panel), GMCP has the highest sharpness values. The sharpness index refers to the predictive distribution concentration [92]. High sharpness indices indicate that the predictive distribution is less dispersed or more concentrated, and therefore high sharpness indices are preferable [6]. The GMCP post-processor improves the sharpness index by 36.64% compared to the MCP and MCPt post-processors.

In terms of the containing ratio (95%CR), the GMCP post-processor outperforms the MCP and MCPt methods. The GMCP improves the 95%CR by 10.29% compared to the MCP

and MCPt, which perform similarly. A proper predictive uncertainty estimation is achieved when the 95%CR comes close to 95%. According to Figure 4 (right panel), the 95%CR obtained by GMCP comes closer to 95%, and with lower variance. The average 95%CR is 93.82% for the GMCP post-processor compared with 85.06% for the MCP or MCPt.

These results show how the boxplots for MCP and MCPt methods overlap. These can indicate that the evaluated reference post-processors that are used with monthly streamflow perform the same in terms of accuracy, sharpness, and reliability. Conversely, the GMCP empirically made the most accurate, reliable, and sharpest predictions for the monthly streamflow of the 12 MOPEX catchments.

Regarding reliability, Figure 5 shows the predictive PQQ plots for the post-processors evaluated through the 12 MOPEX catchments. The PQQ plot indicates the predictive distribution's reliability. According to Figure 5, we stress that the predictive distribution of GMCP (blue line) in most of the evaluated catchments follows the diagonal line in the PQQ, which evidences a reliable predictive uncertainty estimation under different climate conditions. We can also note that the MCP (green curve) and MCPt (red curve) performance is similar for all the evaluated catchments. Furthermore, the PQQ plot for the MCP and MCPt deviates substantially from the 1:1 line in the B4–B9 catchments, indicating some bias. Also, the PQQ plot for the MCP and MCPt in the B11 catchment shows unreliable results, as the predictions are overconfident. In the following subsection, we provide the predictive uncertainty bands of the B11 catchment to explain the poor reliability issue.

3.2. Uncertainty Bands in San Marcos Catchment (B11)

The PQQ plots (Figure 5) evidence that the reference post-processors present reliability problems, while GMCP provides reliable results. The 95% confidence interval of predictive distribution is presented to illustrate these reliability difficulties better. We cannot present the 95% confidence interval for all the catchments and post-processors for space reasons. Given this, we only present the predictive distribution for the San Marcos catchment (B11), a dry catchment, because it clearly shows the reliability problems of the evaluated post-processors.

To illustrate these results, Figure 6 shows the time series of the median and 95% confidence interval of the monthly streamflow forecast at the San Marcos catchment (B11). The GMCP post-processor, which merges clustering with the Gaussian mixture model to deal with heteroscedastic errors, achieves the following verification indices: reliability index = 0.94, sharpness index = 4.44, NSE = 0.94, and the containing ratio (95%CR) = 93.55. Meanwhile, the MCP and MCPt, which perform similarly, have a worse reliability index (metric value = 0.82), sharpness index (metric value = 1.84), NSE (metric value = 0.82), and containing ratio (95%CR) (metric value = 98.16).

In terms of sharpness, the MCP and MCPt methods produce a wider 95% predictive range than the GMCP post-processor (Figures 4 and 6), which manifests as degradation in the sharpness index from 4.44 to 1.84. The widest uncertainty bands produced by the MCP and MCPt confirm the results obtained with the sharpness index (Figure 4, middle panel) and reliability (Figure 5).

Altogether, in the 12 MOPEX catchments, these results show that the GMCP post-processor achieves significant improvements in reliability, sharpness, NSE, and the containing ratio (95%CR). In addition, using the GMCP post-processor for monthly streamflow has an incremental impact on performance, as measured using deterministic and probabilistic verification indices. These results show the robust ability of the GMCP post-processor for better quantifying hydrologic uncertainty and producing enhanced probabilistic streamflow forecasts.

3.3. Influence of the Aridity Index

Figure 7 shows the comparison between the deterministic and probabilistic verification indices for the new GMCP and the two reference post-processors during the validation period (1980–1998) in the 12 MOPEX catchments. Note that the horizontal axis in Figure 7

sorts the catchments from wettest to driest, whereas the vertical axis denotes the post-processor name.

In NSE index terms, and given the purpose of unifying the Figure 7 legend, $|1 - \text{NSE}|$ is shown, where the values close to 0 are the most optimum ones (blue colors in Figure 7). Figure 7 (upper panel) shows the differences in the performance of post-processors in the catchments. For example, the GMCP shows the best performance (blue colors) in most of the evaluated catchments (Figure 7, upper panel). Moreover, the performance in NSE terms for the reference post-processors was generally similar, while the worst performances were obtained in drier catchments (B4, B10, B11, and B12), except for the B5 catchment, which is humid.

In terms of the sharpness index, Figure 7 (middle panel) shows that the lowest sharpness values, which were for the driest catchments (B4, B10, B11 and B12), and the highest values were for the wettest (B2 and B3). In most catchments, the GMCP achieves higher sharpness values than the reference post-processors (MCP and MCPt) (Figure 7, middle panel).

For the 95%CR index, the statistics $|95 - 95\%\text{CR}|$ were calculated, where values close to 0 are preferable and interpreted as the best performance. Similarly, the other verification indices—and as shown in Figure 7 (lower panel)—the GMCP best performed in all the evaluated catchments. Unlike the other verification indices, the 95%CR did not indicate a worsened performance for post-processors in the driest catchments (B4, B10, B11 and B12).

Overall, the results suggest that streamflow forecasts using the GMCP post-processor are better (i.e., NSE and sharpness) than that of the MCP and MCPt methods, particularly in dry catchments. For example, in dry catchments (B4, B10, B11 and B12), the GMCP processor improves the NSE index by 16.66 % compared to the MCP and MCPt methods.

4. Discussion

Predictive uncertainty quantification (PUQ) is essential for supporting effective decision-making and planning for water resources management [93]. In recent years, PUQ has become essential in hydrological predictions [16]. A wide range of methods has been developed to evaluate the predictive uncertainty of the variables of interest. This paper develops an extension of the MCP [14], which merges clusters with Gaussian mixture models to offer an alternative solution to manage heteroscedastic errors. The new method is called the Gaussian mixture clustering post-processor (GMCP). The results of the proposed post-processor were compared to the MCP [14] and the MCPt [35] by applying multiple deterministic and probabilistic verification indices. This research also assesses the GMCP's capacity to estimate the predictive uncertainty of the monthly streamflow under different climate conditions in the 12 MOPEX catchments [70] that are distributed in the SE part of the USA.

Overall, GMCP has shown significant potential in generating more reliable, sharp, and accurate monthly streamflow predictions, especially for dry catchments. Compared to the benchmark methods, GMCP shows more consistency in the validation period than MCP and MCPt (Figure 4). The improvement in the GMCP compared to the MCP and MCPt can be attributed to the procedure used by GMCP to model the dependence structure between observation and forecast (residual error model). GMCP joins the variables via Gaussian mixture models and clusters. Therefore, the Gaussian mixture distribution treats model residuals as three clusters with different means and variances. The Gaussian mixture distribution can capture the peak and the tails of the underlying residual density for all catchments, indicating reliable, sharp, and accurate forecasts. Consequently, this dependence structure of the residual error model faces the assumption of homoscedastic error variance, which provides poor probabilistic predictions. In addition, note that the only difference between the MCP (or MCPt) and GMCP post-processors is the use of clusters and Gaussian mixture models in the GMCP. Hence, the performance improvement must be due to this difference.

Moreover, the MCP and MCPt methods provide similar performances for monthly streamflow predictions regarding the NSE index, reliability, sharpness, and containing ratio (95%CR) (Figures 4–6). The MCPt was designed by Coccia and Todini [35] to improve the reliability and sharpness of predictions, particularly for high flows, and has worked well for flood applications. The MCPt used the truncated Normal distribution (TND) to deal with heteroscedastic errors. In theory, the TDN reduces the standard error of high flows when there is a significant difference between low and high flows, which is the case in flood applications. However, these differences are minor for monthly streamflow, so using the standard Normal distribution or TDN provided similar results.

Figure 5 shows that the PQQ plot for the MCP and MCPt deviates substantially from the 1:1 line in the B4–B9 catchments, indicating some bias. However, the proposed GMCP can obtain unbiased results in the same catchments. One possible explanation is that the MCP uses a linear regression in the Normal transformed space for bias correction, while the GMCP uses three Gaussian mixture models with different means and variances corresponding to the high, middle, and low flow period. Therefore, the bias corrector of GMCP is more robust and flexible than the MCP. As well, Figure 5 depicts that the PQQ plot for the MCP and MCPt in the San Marcos (B11) catchment shows unreliable results, as predictions are overconfident, while GMCP provides reliable results. San Marcos (B11) is a dry catchment with complex residual errors [19]. It is possible that the residual error model of MCP is not enough to represent the complex errors of the San Marcos (B11) catchment. The residual error model of MCP has two assumptions that are undoubtedly inappropriate for the San Marcos (B11) catchment. The MCP assumes homoscedastic errors and a linear relationship between observed and simulated Normal transformed variables. Conversely, the residual error model of GMCP is more complex because of using clustering and Gaussian mixture models.

Our findings in this study confirm the insights of Schaefli et al. [63], namely that using a finite mixture model constitutes a promising solution to residual model errors and to estimate the total modelling uncertainty in hydrological model calibration studies. However, there are two differences between this research and the previous work of Schaefli et al. [63]. First, we used the "post-processing" strategy, where the hydrological model parameters were estimated first using an objective function, followed by a separate estimation of the residual error model parameters. In contrast, Schaefli et al. [63] used the more classical "joint" strategy to estimate all parameters simultaneously using a single likelihood function. Second, we merged the Gaussian mixture model with clusters and used them in the framework of the Model Conditional Processor (MCP) [14]. Likewise, Li et al. [65], who developed the ERRIS post-processor, used a mixture of two Gaussian distributions to represent the residual error model. GMCP and ERRIS have some similarities: (1) both are post-processors of deterministic hydrological models for hydrological uncertainty quantification, (2) both apply a transformation to normalised data, and (3) both use a Gaussian mixture distribution to model residual errors. However, GMCP and ERRIS have some differences. For example, ERRIS uses a linear regression in the transformed space for bias correction, uses an autoregressive model to update hydrological simulation, and is implemented in stages.

We want to discuss some assumptions mentioned in the Materials and Methods in Section 2. First, although GMCP has been conceived to be applied to one single model (point prediction), a multi-model application would be possible. An extension of the GMCP consists of a matrix of predictions and various deterministic models (one column for each model), yet here we study the simpler scalar version of the model. Second, there are possible ways to advance towards the application of GMCP in a non-stationary context. The simplest option is using a deterministic model for non-stationarity [71,72]. We also suggest considering a deterministic model with time-varying (perhaps seasonal) parameters, under the assumption that the uncertainty of the model for non-stationarity is represented by a stationary distribution. In addition, we recommend the use of data assimilation to update hydrological predictions [94]. Third, in the GMCP, we used Gaussian mixture distributions

and fixed the number of mixture components (clusters) to three—corresponding to the high, middle, and low flow period, which are typical of monthly streamflow. This practical choice is based on a priori information about the sources and behaviour of the residual error model. Therefore, identifying the number of clusters is purely heuristic accounting for a priori knowledge about the total error model. Fourth, in the GMCP, any probabilistic prediction is primarily based on the conditions monitored during the considered observation period only, and thus particular care should be used when extrapolation to out-of-sample conditions.

This research confirms the importance of using multiple independent verification indices to assess hydrological post-processors. For example, if one considers the containing ratio (95%CR) verification index alone, all post-processors yield comparable performances, and there is no argument for selecting any of them. Nonetheless, once the sharpness index and reliability index are considered explicitly, the GMCP post-processor can be recommended for significantly better sharpness and reliability than the MCP and MCPt. These results align with Woldemeskel et al. [41], who showed that evaluating the predictive uncertainty with a single metric can lead to suboptimum conclusions.

Moreover, examining and evaluating hydrological post-processors in catchments with different climate and hydrological conditions ensures suitable comparisons and helps to generalise the obtained results [47,95]. Furthermore, the diverse climate conditions of catchments analysed allow us to deduce functional relationships between climatic indices and the post-processors' performance. This research attempted to establish a relation between the Aridity index and the post-processors' performance (Figure 7). In most dry catchments, the MCP and MCPt perform relatively worse, especially in terms of the sharpness and NSE index. This result is because streamflow data for dry catchments contain too many days with low flow (defined as flow below 2% of the mean flow [12]). Thus, dry catchments require more complex residual error modelling methods [64]. Our findings agree with Ye et al. [19], who found that the GLMPP post-processor [26] could not improve the predictions or reduce uncertainty in the same dry MOPEX catchments.

In this study, all post-processors provide a clear improvement in hydrological predictions. Post-processing usually leads to better performance verification indices than deterministic hydrological predictions alone because post-processing works directly to correct the errors in the model outputs [19]. Accordingly, Farmer and Vogel [22] stated that the prudent management of environmental resources requires probabilistic predictions, which offer the potential to quantify predictive uncertainty, and can avoid the false sense of security associated with point predictions [16]. Generally speaking, predictive distribution explicitly represents the system's uncertainty, and it can, therefore, perform risk management in a more informed manner. In addition, the probabilistic approach can be put to further use for process-based deterministic hydrological modelling and by coupling it with a hydrological post-processor to convert deterministic predictions into probabilistic predictions. Probabilistic predictions offer an opportunity to improve the operational planning and management of water resources.

A promising improvement is to extend the GMCP post-processor in future work using a multi-model or a chain of hydrological models. Also, it can be interesting to validate the GMCP using daily and hourly data and couple the GMCP method with data assimilation to update the states of hydrological models. Bourgin et al. [94] recommended using data assimilation and post-processing in forecasting because data assimilation strongly impacts forecast accuracy, while post-processing strongly impacts forecast reliability. Besides, evaluating the GMCP post-processor in many catchments is beneficial for establishing its robustness.

Another area for further investigation is overcoming the data transformation. In hydrology, data transformation is a popular approach to reduce the heteroscedasticity of the error model because these approaches are simple to implement and can give satisfactory results in hydrological modelling [14,41,51,65]. However, Schaefli et al. [63], Brown and Seo [42], and others indicated that this approach is questionable. A detailed discussion of the implications of data transformation is beyond the context of this paper. Nevertheless,

we recommend reading the work of Hunter et al. [96], which established the detrimental impact of calibrating hydrological parameters in the real space and calibrating the error model parameters in the transformed space using post-processing methods on the quality of probabilistic predictions. Moreover, a future possible study could be to extend the GMCP post-processor using a link function to avoid the Normal quantile transformation, especially the link function, which has provided promising results in the context of the Generalized Linear Model. Finally, another improvement can be the selection of the number of GMCP clusters using unsupervised learning, for example, by using cluster indicators [97].

5. Conclusions and Summary

Considering that predictive uncertainty is crucial for providing reliable, sharp, and accurate probabilistic streamflow predictions, the Model Conditional Processor (MCP) [14] is a well-known method for quantifying predictive uncertainty by providing a posterior distribution conditioned on the deterministic model forecast. This study develops an extension of the MCP [14], which merges clustering with the Gaussian mixture model to offer an alternative solution to manage heteroscedastic errors. The new method is called the Gaussian mixture clustering post-processor (GMCP). The results of the proposed post-processor were compared to the MCP [14] and the MCPt [35] by applying multiple deterministic and probabilistic verification indices. This research also assesses the GMCP's capacity to estimate the predictive uncertainty of the monthly streamflow under different climate conditions in the 12 MOPEX catchments [70] distributed in the SE part of the USA. The summary of the most important empirical findings based on the detailed analysis of the results are as follows:

1. In general, all three post-processors showed promising results. However, the GMCP post-processor has shown significant potential in generating more reliable, sharp, and accurate monthly streamflow predictions than the MCP and MCPt methods, especially in dry catchments.
2. The MCP and MCPt methods provided similar performances for monthly streamflow predictions regarding the NSE index, reliability, sharpness, and containing ratio (95%CR).
3. The MCP and MCPt showed a better performance in wet catchments than in dry catchments.

Overall, when used for post-processing monthly predictions, the GMCP method provides an opportunity to improve forecast performance further than is possible using the MCP and MCPt methods, especially in dry catchments. In addition, it is worth mentioning that incorporating clusters and Gaussian mixture models into the Model Conditional Processor framework constitutes a promising solution to handle heteroscedastic errors in monthly streamflow, therefore moving towards a more realistic monthly hydrological prediction to support effective decision-making in planning and managing water resources.

Author Contributions: Conceptualization, F.F., M.R.H.-L. and J.R.-C.; methodology and software, M.R.H.-L. and J.R.-C.; formal analysis, F.F. and J.R.-C.; data curation and visualization, C.J.G.-T., M.R.H.-L. and J.R.-C.; writing—original draft preparation, J.R.-C.; writing—review and editing, C.P.S., F.F. and J.R.-C.; supervision and funding acquisition, F.F. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All data for the US catchments are taken from MOPEX dataset, which are available at http://www.nws.noaa.gov/ohd/mopex/mo_datasets.htm (accessed on 15 May 2018). The code used to generate the results presented in this study can be found on GitHub (https://github.com/jhonrom/GMCP.git) (accessed on 25 January 2022).

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

References

1. Stakhiv, E.; Stewart, B. Needs for Climate Information in Support of Decision-Making in the Water Sector. *Procedia Environ. Sci.* **2010**, *1*, 102–119. [CrossRef]
2. Hamilton, S.H.; Fu, B.; Guillaume, J.H.; Badham, J.; Elsawah, S.; Gober, P.; Hunt, R.J.; Iwanaga, T.; Jakeman, A.J.; Ames, D.P.; et al. A framework for characterising and evaluating the effectiveness of environmental modelling. *Environ. Model. Softw.* **2019**, *118*, 83–98. [CrossRef]
3. Chang, F.J.; Guo, S. Advances in Hydrologic Forecasts and Water Resources Management. *Water* **2020**, *12*, 1819. [CrossRef]
4. Kavetski, D.; Kuczera, G.; Franks, S.W. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resour. Res.* **2006**, *42*, W03408. [CrossRef]
5. Gan, Y.; Liang, X.-Z.; Duan, Q.; Ye, A.; Di, Z.; Hong, Y.; Li, J. A systematic assessment and reduction of parametric uncertainties for a distributed hydrological model. *J. Hydrol.* **2018**, *564*, 697–711. [CrossRef]
6. Renard, B.; Kavetski, D.; Kuczera, G.; Thyer, M.; Franks, S.W. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resour. Res.* **2010**, *46*, W05521. [CrossRef]
7. Bulygina, N.; Gupta, H. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. *Water Resour. Res.* **2009**, *45*, W00B13. [CrossRef]
8. Clark, M.P.; Slater, A.G.; Rupp, D.E.; Woods, R.A.; Vrugt, J.A.; Gupta, H.V.; Wagener, T.; Hay, L.E. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resour. Res.* **2008**, *44*, W00B02. [CrossRef]
9. Reichle, R.H.; Mclaughlin, D.B.; Entekhabi, D. Hydrologic Data Assimilation with the Ensemble Kalman Filter. *American Meteorol. Soc.* **2002**, *130*, 103–114. [CrossRef]
10. Clark, M.P.; Kavetski, D. Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. *Water Resour. Res.* **2010**, *46*, W10510. [CrossRef]
11. Reichert, P. Conceptual and Practical Aspects of Quantifying Uncertainty in Environmental Modelling and Decision Support. 2012. Available online: https://www.semanticscholar.org/paper/Conceptual-and-Practical-Aspects-of-Quantifying-in-Reichert/4dbd0397c9cb925cff1eea445e8d18428ef4a95a (accessed on 16 July 2019).
12. McInerney, D.; Thyer, M.; Kavetski, D.; Bennett, B.; Lerat, J.; Gibbs, M.; Kuczera, G. A simplified approach to produce probabilistic hydrological model predictions. *Environ. Model. Softw.* **2018**, *109*, 306–314. [CrossRef]
13. Krzysztofowicz, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **1999**, *35*, 2739–2750. [CrossRef]
14. Todini, E. A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manag.* **2008**, *6*, 123–137. [CrossRef]
15. Loucks, D.P.; van Beek, E. *Water Resource Systems Planning and Management*; Springer International Publishing: Cham, Switzerland, 2017.
16. Todini, E. Paradigmatic changes required in water resources management to benefit from probabilistic forecasts. *Water Secur.* **2018**, *3*, 9–17. [CrossRef]
17. Refsgaard, J.C.; van der Sluijs, J.P.; Højberg, A.L.; Vanrolleghem, P.A. Uncertainty in the environmental modelling process—A framework and guidance. *Environ. Model. Softw.* **2007**, *22*, 1543–1556. [CrossRef]
18. Prieto, C.; le Vine, N.; Kavetski, D.; García, E.; Medina, R. Flow Prediction in Ungauged Catchments Using Probabilistic Random Forests Regionalization and New Statistical Adequacy Tests. *Water Resour. Res.* **2019**, *55*, 4364–4392. [CrossRef]
19. Ye, A.; Duan, Q.; Yuan, X.; Wood, E.F.; Schaake, J. Hydrologic post-processing of MOPEX streamflow simulations. *J. Hydrol.* **2014**, *508*, 147–156. [CrossRef]
20. Hopson, T.M.; Wood, A.; Weerts, A.H. Motivation and overview of hydrological ensemble post-processing. In *Handbook of Hydrometeorological Ensemble Forecasting*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 783–793.
21. Montanari, A.; Koutsoyiannis, D. A blueprint for process-based modeling of uncertain hydrological systems. *Water Resour. Res.* **2012**, *48*, W09555. [CrossRef]
22. Farmer, W.H.; Vogel, R.M. On the deterministic and stochastic use of hydrologic models. *Water Resour. Res.* **2016**, *52*, 5619–5633. [CrossRef]
23. Montanari, A.; Brath, A. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.* **2004**, *40*, W01106. [CrossRef]
24. Montanari, A.; Grossi, G. Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resour. Res.* **2008**, *44*, W00B08. [CrossRef]
25. Wang, Q.J.; Robertson, D.E.; Chiew, F.H.S. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resour. Res.* **2009**, *45*, W05407. [CrossRef]

26. Zhao, L.; Duan, Q.; Schaake, J.; Ye, A.; Xia, J. A hydrologic post-processor for ensemble streamflow predictions. *Adv. Geosci.* **2011**, *29*, 51–59. [CrossRef]

27. McInerney, D.; Thyer, M.; Kavetski, D.; Lerat, J.; Kuczera, G. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resour. Res.* **2017**, *53*, 2199–2239. [CrossRef]

28. Schoups, G.; Vrugt, J.A. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.* **2010**, *46*, W10531. [CrossRef]

29. Smith, T.; Marshall, L.; Sharma, A. Modeling residual hydrologic errors with Bayesian inference. *J. Hydrol.* **2015**, *528*, 29–37. Available online: https://www.sciencedirect.com/science/article/pii/S0022169415004011 (accessed on 3 May 2017). [CrossRef]

30. Sorooshian, S.; Dracup, J.A. Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resour. Res.* **1980**, *16*, 430–442. [CrossRef]

31. Van Der Waerden, B. Order tests for two-sample problem and their power I. *Indag. Math.* **1952**, *55*, 453–458. [CrossRef]

32. Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc.* **1964**, *26*, 211–252. Available online: https://www.semanticscholar.org/paper/An-Analysis-of-Transformations-Box-Cox/6e820cf11712b9041bb625634612a535476f0960 (accessed on 24 May 2019). [CrossRef]

33. 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*, W05514. [CrossRef]

34. Krzysztofowicz, R.; Kelly, K.S. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* **2000**, *36*, 3265–3277. [CrossRef]

35. 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]

36. 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]

37. Tyralis, H.; Papacharalampous, G. Quantile-Based Hydrological Modelling. *Water* **2021**, *13*, 3420. [CrossRef]

38. 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]

39. Darbandsari, P.; Coulibaly, P. Inter-Comparison of Different Bayesian Model Averaging Modifications in Streamflow Simulation. *Water* **2019**, *11*, 1707. [CrossRef]

40. Evin, G.; Thyer, M.; Kavetski, D.; McInerney, D.; Kuczera, G. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resour. Res.* **2014**, *50*, 2350–2375. [CrossRef]

41. Woldemeskel, F.; McInerney, D.; Lerat, J.; Thyer, M.; Kavetski, D.; Shin, D.; Tuteja, N.; Kuczera, G. Evaluating post-processing approaches for monthly and seasonal streamflow forecasts. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6257–6278. [CrossRef]

42. 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]

43. Solomatine, D.P.; Shrestha, D.L. A novel method to estimate model uncertainty using machine learning techniques. *Water Resour. Res.* **2009**, *45*, W00B11. [CrossRef]

44. López, 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]

45. Sikorska, A.E.; Montanari, A.; Koutsoyiannis, D. Estimating the Uncertainty of Hydrological Predictions through Data-Driven Resampling Techniques. *J. Hydrol. Eng.* **2015**, *20*, A4014009. [CrossRef]

46. Ehlers, L.B.; Wani, O.; Koch, J.; Sonnenborg, T.O.; Refsgaard, J.C. Using a simple post-processor to predict residual uncertainty for multiple hydrological model outputs. *Adv. Water Resour.* **2019**, *129*, 16–30. [CrossRef]

47. Tyralis, H.; Papacharalampous, G.; Burnetas, A.; Langousis, A. Hydrological post-processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS. *J. Hydrol.* **2019**, *577*, 123957. [CrossRef]

48. Papacharalampous, G.; Tyralis, H.; Langousis, A.; Jayawardena, A.W.; Sivakumar, B.; Mamassis, N.; Montanari, A.; Koutsoyiannis, D. Probabilistic Hydrological Post-Processing at Scale: Why and How to Apply Machine-Learning Quantile Regression Algorithms. *Water* **2019**, *11*, 2126. [CrossRef]

49. Schefzik, R.; Thorarinsdottir, T.L.; Gneiting, T. Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling. *Stat. Sci.* **2013**, *28*, 616–640. [CrossRef]

50. Madadgar, S.; Moradkhani, H. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resour. Res.* **2014**, *50*, 9586–9603. [CrossRef]

51. Klein, B.; Meissner, D.; Kobialka, H.-U.; Reggiani, P. Predictive Uncertainty Estimation of Hydrological Multi-Model Ensembles Using Pair-Copula Construction. *Water* **2016**, *8*, 125. [CrossRef]

52. Li, W.; Duan, Q.; Miao, C.; Ye, A.; Gong, W.; Di, Z. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdiscip. Rev. Water* **2017**, *4*, e1246. [CrossRef]

53. Moges, E.; Demissie, Y.; Larsen, L.; Yassin, F. Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water* **2020**, *13*, 28. [CrossRef]

54. Matott, L.S.; Babendreier, J.E.; Purucker, S.T. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resour. Res.* **2009**, *45*, 6421. [CrossRef]

55. Reggiani, P.; Coccia, G.; Mukhopadhyay, B. Predictive Uncertainty Estimation on a Precipitation and Temperature Reanalysis Ensemble for Shigar Basin, Central Karakoram. *Water* **2016**, *8*, 263. [CrossRef]

56. Barbetta, S.; Coccia, G.; Moramarco, T.; Todini, E. Case Study: A Real-Time Flood Forecasting System with Predictive Uncertainty Estimation for the Godavari River, India. *Water* **2016**, *8*, 463. [CrossRef]

57. Biondi, D.; Todini, E. Comparing Hydrological Postprocessors Including Ensemble Predictions into Full Predictive Probability Distribution of Streamflow. *Water Resour. Res.* **2018**, *54*, 9860–9882. [CrossRef]

58. Massari, C.; Maggioni, V.; Barbetta, S.; Brocca, L.; Ciabatta, L.; Camici, S.; Moramarco, T.; Coccia, G.; Todini, E. Complementing near-real time satellite rainfall products with satellite soil moisture-derived rainfall through a Bayesian Inversion approach. *J. Hydrol.* **2019**, *573*, 341–351. [CrossRef]

59. Parviz, L.; Rasouli, K. Development of Precipitation Forecast Model Based on Artificial Intelligence and Subseasonal Clustering. *J. Hydrol. Eng.* **2019**, *24*, 04019053. [CrossRef]

60. Yu, Y.; Shao, Q.; Lin, Z. Regionalization study of maximum daily temperature based on grid data by an objective hybrid clustering approach. *J. Hydrol.* **2018**, *564*, 149–163. [CrossRef]

61. Basu, B.; Srinivas, V.V. Regional flood frequency analysis using kernel-based fuzzy clustering approach. *Water Resour. Res.* **2014**, *50*, 3295–3316. [CrossRef]

62. Zhang, H.; Huang, G.H.; Wang, D.; Zhang, X. Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering. *Adv. Water Resour.* **2011**, *34*, 1292–1303. [CrossRef]

63. Schaefli, B.; Talamba, D.B.; Musy, A. Quantifying hydrological modeling errors through a mixture of normal distributions. *J. Hydrol.* **2007**, *332*, 303–315. [CrossRef]

64. Smith, T.; Sharma, A.; Marshall, L.; Mehrotra, R.; Sisson, S. Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. *Water Resour. Res.* **2010**, *46*, W12551. [CrossRef]

65. Li, M.; Wang, Q.J.; Bennett, J.C.; Robertson, D.E. Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3561–3579. [CrossRef]

66. Feng, K.; Zhou, J.; Liu, Y.; Lu, C.; He, Z. Hydrological Uncertainty Processor (HUP) with Estimation of the Marginal Distribution by a Gaussian Mixture Model. *Water Resour. Manag.* **2019**, *33*, 2975–2990. [CrossRef]

67. Yang, X.; Zhou, J.; Fang, W.; Wang, Y. An Ensemble Flow Forecast Method Based on Autoregressive Model and Hydrological Uncertainty Processer. *Water* **2020**, *12*, 3138. [CrossRef]

68. Kim, K.H.; Yun, S.T.; Park, S.S.; Joo, Y.; Kim, T.S. Model-based clustering of hydrochemical data to demarcate natural versus human impacts on bedrock groundwater quality in rural areas, South Korea. *J. Hydrol.* **2014**, *519*, 626–636. [CrossRef]

69. Duan, Q.; Schaake, J.; Andréassian, V.; Franks, S.; Goteti, G.; Gupta, H.; Gusev, Y.; Habets, F.; Hall, A.; Hay, L.; et al. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.* **2006**, *320*, 3–17. [CrossRef]

70. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria. 2013. Available online: Ftp: //ftp.uvigo.es/CRAN/web/packages/dplR/vignettes/intro-dplR.pdf (accessed on 16 April 2019).

71. Pathiraja, S.; Marshall, L.; Sharma, A.; Moradkhani, H. Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. *Adv. Water Resour.* **2016**, *94*, 103–119. [CrossRef]

72. Deb, P.; Kiem, A.S. Evaluation of rainfall–runoff model performance under non-stationary hydroclimatic conditions. *Hydrol. Sci. J.* **2020**, *65*, 1667–1684. [CrossRef]

73. Simonoff, J.S. *Smoothing Methods in Statistics*; Springer: New York, NY, USA, 1996.

74. Duong, T. Ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *J. Stat. Softw.* **2007**, *21*, 1–16. [CrossRef]

75. Wolfe, J.H. Pattern clustering by multivariate mixture analysis. *Multivariate Behav. Res.* **1970**, *5*, 329–350. [CrossRef]

76. Boehmke, B.; Greenwell, B.M. *Hands-On Machine Learning with R*, 1st ed.; Chapman and Hall: London, UK; CRC: Boca Raton, FL, USA, 2019.

77. Banfield, J.D.; Raftery, A.E. Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics* **1993**, *49*, 821. [CrossRef]

78. Fraley, C.; Raftery, A.E. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611–631. [CrossRef]

79. James, L.F.; Priebe, C.E.; Marchette, D.J. Consistent estimation of mixture complexity. *Ann. Stat.* **2001**, *29*, 1281–1296. [CrossRef]

80. McLachlan, G.J.; Peel, D. *Finite Mixture Models*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2000.

81. Melnykov, V.; Maitra, R. Finite mixture models and model-based clustering. *Stat. Surv.* **2010**, *4*, 80–116. [CrossRef]

82. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*; Wiley-Interscience: Hoboken, NJ, USA, 2008.

83. Zhang, W.; Di, Y. Model-based clustering with measurement or estimation errors. *Genes* **2020**, *11*, 185. [CrossRef] [PubMed]

84. Franz, K.J.; Hogue, T.S. Evaluating uncertainty estimates in hydrologic models: Borrowing measures from the forecast verification community. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3367–3382. [CrossRef]

85. Perrin, C.; Michel, C.; Andréassian, V. Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* **2003**, *279*, 275–289. [CrossRef]

86. Laio, F.; Tamea, S. Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1267–1277. [CrossRef]

87. Thyer, M.; Renard, B.; Kavetski, D.; Kuczera, G.; Franks, S.W.; Srikanthan, S. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resour. Res.* **2009**, *45*, W00B14. [CrossRef]

88. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]

89. Kavetski, D.; Clark, M.P. Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. *Water Resour. Res.* **2010**, *46*, 2009WR008896. [CrossRef]

90. Gupta, H.V.; Perrin, C.; Blöschl, G.; Montanari, A.; Kumar, R.; Clark, M.; Andréassian, V. Large-sample hydrology: A need to balance depth with breadth. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 463–477. [CrossRef]

91. Martinez, G.F.; Gupta, H.V. Toward improved identification of hydrological models: A diagnostic evaluation of the 'abcd' monthly water balance model for the conterminous United States. *Water Resour. Res.* **2010**, *46*, W08507. [CrossRef]

92. 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]

93. Tolson, B.A.; Shoemaker, C.A. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resour. Res.* **2008**, *44*, 4411. [CrossRef]

94. Bourgin, F.; Ramos, M.H.; Thirel, G.; Andréassian, V. Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting. *J. Hydrol.* **2014**, *519*, 2775–2784. [CrossRef]

95. Papacharalampous, G.; Tyralis, H.; Koutsoyiannis, D.; Montanari, A. Quantification of predictive uncertainty in hydrological modelling by harnessing the wisdom of the crowd: A large-sample experiment at monthly timescale. *Adv. Water Resour.* **2020**, *136*, 103470. [CrossRef]

96. Hunter, J.; Thyer, M.; McInerney, D.; Kavetski, D. Achieving high-quality probabilistic predictions from hydrological models calibrated with a wide range of objective functions. *J. Hydrol.* **2021**, *603*, 126578. [CrossRef]

97. Breaban, M.; Luchian, H. A unifying criterion for unsupervised clustering and feature selection. *Pattern Recognit.* **2011**, *44*, 854–865. [CrossRef]