Universitat Politècnica de València

Departamento de Ingeniería Hidráulica y Medio Ambiente

Programa de Doctorado de Ingeniería del Agua y Medioambiental

Improving hydrological post-processing for assessing the conditional predictive uncertainty of monthly streamflows

AUTHOR: JONATHAN ROMERO CUÉLLAR

SUPERVISOR: DR. FÉLIX FRANCÉS GARCÍA

VALENCIA, NOVEMBER 2019

To my family,

One thing I do know: I was blind,
And now I see.
Joan 9:25

Remember the LORD in everything you do,
And he will show you the right way.
Proverbs 3:6

Those who become wise are happy;
Wisdom will give them life.
Proverbs 3:18

ACKNOWLEDGEMENTS

First, I thank God for guiding me through the right way in my life.

I would like to thank the Gobernación del Huila Scholarship Program No. 677 (Colombia) for providing the financial support for my PhD research.

I would like to express my deepest gratitude to my PhD mentor, Prof. Félix Francés for giving me the opportunity to work with you. It has been a great learning experience for me.

I also thank Professor Giada Adelfio for housing me during a research visit to the Department of Economics, Business and Statistics at the Universitá degli Studi di Palermo, Italy. You taught me to love the probabilistic framework and non-parametric statistics.

I am grateful to Qingyun Duan for information of the MOPEX experiment.

I also thank my colleagues at the Research Group of Hydrological and Environmental Modelling (GIMHA) and my friends' role in my life, especially my faith family (D13).

I am really thankful my picola anima Sabribella for her great role in my life. Thanks for being there to celebrate my happiness and also to be the biggest support during those difficult times.

Finalmente, quiero dar un especial reconocimiento a mi familia, quienes son los verdaderos merecedores de este trabajo. Primero, a mi mamá por enseñarme con su ejemplo los valores de responsabilidad y cumplimiento, a mi papá por enseñarme a valorar y tratar a las personas con amor y compasión y mis hermanos por estar siempre conmigo, aunque sea a la distancia (Nana y Carlitos), finalmente mis sobrinas Anyelita y Ami por aguantar tanto tiempo sin su tío preferido.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The predictive uncertainty quantification in monthly streamflows is crucial to make reliable hydrological predictions that help and support decision-making in water resources management. Hydrological post-processing methods are suitable tools to estimate the predictive uncertainty of deterministic streamflow predictions (hydrological model outputs). In general, this thesis focuses on improving hydrological post-processing methods for assessing the conditional predictive uncertainty of monthly streamflows. This thesis deal with two issues of the hydrological post-processing scheme i) the heteroscedasticity problem and ii) the intractable likelihood problem. Mainly, this thesis includes three specific aims. First and relate to the heteroscedasticity problem, we develop and evaluate a new post-processing approach, called GMM post-processor, which is based on the Bayesian joint probability modelling approach and the Gaussian mixture models. Besides, we compare the performance of the proposed post-processor with the well-known exiting post-processors for monthly streamflows across 12 MOPEX catchments. From this aim (chapter 2), we find that the GMM post-processor is the best suited for estimating the conditional predictive uncertainty of monthly streamflows, especially for dry catchments.

Secondly, we introduce a method to quantify the conditional predictive uncertainty in hydrological post-processing contexts when it is cumbersome to calculate the likelihood (intractable likelihood). Sometimes, it can be challenging to estimate the likelihood itself in hydrological modelling, especially working with complex models or with ungauged catchments. Therefore, we propose the ABC post-processor that exchanges the requirement of calculating the likelihood function by the use of some sufficient summary statistics and synthetic datasets. With this aim in mind (chapter 3), we prove that the conditional predictive distribution is similarly produced by the exact predictive (MCMC post-processor) or the approximate predictive (ABC post-processor), qualitatively speaking. This finding is significant because dealing with scarce information is a common condition in hydrological studies.

Finally, we apply the ABC post-processing method to estimate the uncertainty of streamflow statistics obtained from climate change projections, such as a particular case of intractable likelihood problem. From this specific objective (chapter 4), we find that the ABC post-processor approach: 1) offers more reliable projections than 14 climate models (without post-processing); 2) concerning the best climate models during the baseline period, produces more realistic uncertainty bands than the classical multi-model ensemble approach.

RESUMEN

La cuantificación de la incertidumbre predictiva es de vital importancia para producir predicciones hidrológicas confiables que soporten y apoyen la toma de decisiones en el marco de la gestión de los recursos hídricos. Los post-procesadores hidrológicos son herramientas adecuadas para estimar la incertidumbre predictiva de las predicciones hidrológicas (salidas del modelo hidrológico). El objetivo general de esta tesis es mejorar los métodos de post-procesamiento hidrológico para estimar la incertidumbre predictiva de caudales mensuales. Esta tesis pretende resolver dos problemas del post-procesamiento hidrológico: i) la heterocedasticidad y ii) la función de verosimilitud intratable. Los objetivos específicos de esta tesis son tres. Primero y relacionado con la heterocedasticidad, se propone y evalúa un nuevo método de post-procesamiento llamado GMM post-processor que consiste en la combinación del esquema de modelado de probabilidad Bayesiana conjunta y la mezcla de Gaussianas múltiples. Además, se comparó el desempeño del post-procesador propuesto con otros métodos tradicionales y bien aceptados en caudales mensuales a través de las doce cuencas hidrográficas del proyecto MOPEX. A partir de este objetivo (capitulo 2), encontramos que GMM post-processor es el mejor para estimar la incertidumbre predictiva de caudales mensuales, especialmente en cuencas de clima seco.

Segundo, se propone un método para cuantificar la incertidumbre predictiva en el contexto de post-procesamiento hidrológico cuando sea difícil calcular la función de verosimilitud (función de verosimilitud intratable). Algunas veces en modelamiento hidrológico es difícil calcular la función de verosimilitud, por ejemplo, cuando se trabaja con modelos complejos o en escenarios de escasa información como en cuencas no aforadas. Por lo tanto, se propone el ABC post-processor que intercambia la estimación de la función de verosimilitud por el uso de resúmenes estadísticos y datos simulados. De este objetivo específico (capitulo 3), se demuestra que la distribución predictiva estimada por un método exacto (MCMC post-processor) o por un método aproximado (ABC post-processor) es similar. Este resultado es importante porque trabajar con escasa información es una característica común en los estudios hidrológicos.

Finalmente, se aplica el ABC post-processor para estimar la incertidumbre de los estadísticos de los caudales obtenidos desde las proyecciones de cambio climático, como un caso particular de un problema de función de verosimilitud intratable. De este objetivo específico (capitulo 4), encontramos que el ABC post-processor ofrece proyecciones de cambio climático más confiables que los 14 modelos climáticos (sin post-procesamiento). De igual forma, ABC post-processor produce bandas de incertidumbre más realista para los estadísticos de los caudales que el método clásico de múltiples conjuntos (ensamble).

Resum

La quantificació de la incertesa predictiva és de vital importància per a produir prediccions hidrològiques confiables que suporten i recolzen la presa de decisions en el marc de la gestió dels recursos hídrics. Els post-processadors hidrològics són eines adequades per a estimar la incertesa predictiva de les prediccions hidrològiques (eixides del model hidrològic). L'objectiu general d'aquesta tesi és millorar els mètodes de post-processament hidrològic per a estimar la incertesa predictiva de cabals mensuals. Els objectius específics d'aquesta tesi són tres. Primer, es proposa i avalua un nou mètode de post-processament anomenat GMM post-processor que consisteix en la combinació de l'esquema de modelatge de probabilitat Bayesiana conjunta i la barreja de Gaussianes múltiples. A més, es compara l'acompliment del post-processador proposat amb altres mètodes tradicionals i ben acceptats en cabals mensuals a través de les dotze conques hidrogràfiques del projecte MOPEX. A partir d'aquest objectiu (capítol 2), trobem que GMM post-processor és el millor per a estimar la incertesa predictiva de cabals mensuals, especialment en conques de clima sec.

En segon lloc, es proposa un mètode per a quantificar la incertesa predictiva en el context de post-processament hidrològic quan siga difícil calcular la funció de versemblança (funció de versemblança intractable). Algunes vegades en modelació hidrològica és difícil calcular la funció de versemblança, per exemple, quan es treballa amb models complexos o amb escenaris d'escassa informació com a conques no aforades. Per tant, es proposa l'ABC post-processor que intercanvia l'estimació de la funció de versemblança per l'ús de resums estadístics i dades simulades. D'aquest objectiu específic (capítol 3), es demostra que la distribució predictiva estimada per un mètode exacte (MCMC post-processor) o per un mètode aproximat (ABC post-processor) és similar. Aquest resultat és important perquè treballar amb escassa informació és una característica comuna als estudis hidrològics.

Finalment, s'aplica l'ABC post-processor per a estimar la incertesa dels estadístics dels cabals obtinguts des de les projeccions de canvi climàtic. D'aquest objectiu específic (capítol 4), trobem que l'ABC post-processor ofereix projeccions de canvi climàtic més confiables que els 14 models climàtics (sense post-processament). D'igual forma, ABC post-processor produeix bandes d'incertesa més realistes per als estadístics dels cabals que el mètode clàssic d'assemble.

CHAPTER 1. Introduction

1.1 Probabilistic uncertainty quantification

Hydrological predictions provide crucial supporting information for effective decision making in the water sector, such as: planning for new investments (irrigation systems, water supply, and reservoirs), flood emergency response, water allocation, ecological issues, drought risk management, operation and monitoring of existing systems to adapt new conditions (climate change, land-use change, population growth, etc.) and introduce new efficient technology (wastewater reuse, biotechnology, desalting, solar energy) (Stakhiv and Stewart, 2010). However, predictions are affected by different sources of uncertainty, such as: observed data uncertainty, parametric uncertainty, structural uncertainty, initial condition uncertainty, numerical solution uncertainty and non-deterministic behaviour of a system (Reichert, 2012; Renard et al., 2010). Specifically, there are three critical sources of errors. First, the rainfall observations are affected by sampling errors, which arise from an incomplete sampling of the spatially and temporally distributed random fields (Kuczera et al., 2006). Secondly, the streamflow observations are subject to rating curve errors, which is most often stage-discharge rating curve models that may be imprecise and/or biased (McMillan et al., 2017) and the observed data are also affected by measurement errors. Thirdly, given that the hydrological models are always a simplified representation of reality, they produce a discrepancy between simulated and observed variables. This discrepancy is called “Model Structural Error” and may consist of the inadequate selection of process formulation, model variables and spatial and temporal resolution of the model (Reichert and Mieleitner, 2009).

For all these reasons, predictions are always uncertain because models can never correctly represent a natural system. Therefore, an estimate of the uncertainty of these predictions is required. Different motives demonstrate the importance of uncertainty quantification (UQ). Montanari (2011), Reichert et al. (2015), Vogel (2017) and Kavetski (2019) claimed that UQ is a hot topic and is a still research challenge -especially because the UQ methods have not been standardised (Montanari and Koutsoyiannis, 2012; van Oijen, 2017; Wagener and Gupta, 2005). Besides, Butts et al. (2004), Liu and Gupta (2007), Reichert et al. (2015) and Kavetski (2019) stated that UQ is vital for making better-informed decisions. Schoups et al. (2008) expressed that UQ is also useful for advancing towards reliable measurement systems, modelling comparison and selection. Moreover, Schoups and Vrugt (2010) stated that hydrological predictions without UQ are impractical tools to plan and operate water resources systems. UQ is also considered a desirable scientific practice (Refsgaard et al., 2007). Montanari (2011) mentioned that UQ plays a fundamental role in the learning process, while Schoups and Vrugt (2010) and Wang et al. (2016) supported that UQ is an essential tool for improving model performances. Finally, McMillan et al. (2017) argued that including UQ as a standard part of related water management applications can lead to cost savings and increase the robustness of decisions.

Broadly speaking, uncertainty can be defined as an attribute of information (Zadeh, 2005); more specifically, it is an inherent property of the hydrological process (Montanari, 2011). The uncertainty quantification (UQ) seeks to characterise the entire set of possible outcomes, together with their associated probabilities of occurrence (Loucks and van Beek, 2017). Uncertainty can be classified in aleatory and epistemic. Aleatory uncertainty refers to the inherent variability as well as the randomness of underlying phenomena, while epistemic uncertainty concerns the incomplete knowledge of the modelled system. Both types of uncertainty are manageable using probabilistic methods, of which the probability theory is the base (Kavetski, 2019; Montanari, 2011).

Many techniques for estimating these uncertainties have been proposed, including the data assimilation for initial condition errors (Moradkhani et al., 2005a), many approaches for model structural errors (Butts et al., 2004; Clark et al., 2008; Hoeting et al., 1999; Vrugt and Robinson, 2007), several Bayesian schemes for parametric, input and output uncertainty (Blazkova and Beven, 2009; Honti et al., 2013; Kavetski et al., 2018, 2011; Renard et al., 2010; Schoups and Vrugt, 2010; Thyer et al., 2009; Vrugt and Sadegh, 2013), ensembles for chaotic behaviour (Stainforth et al., 2007; Tebaldi and Knutti, 2007; van der Linden and Mitchell, 2009), some guidelines for numerical errors (Clark and Kavetski, 2010; Kavetski and Clark, 2010) and various statistical post-processing approaches (W. Li et al., 2017).

A hydrological post-processor is a statistical technique used to improve typical deterministic forecasts by relating hydrological model outputs to observations (Ye et al., 2014). Since the hydrological modelling process has many sources of uncertainties, post-processing is required to characterise these uncertainties and remove systemic bias in the predicting process (Hopson et al., 2019). Post-processors dilute errors from model inputs and outputs, model parameters, model initial and boundary conditions and model structures (Buizza, 2018; Ye et al., 2014). Hydrological post-processing methods mostly follow the Model Output Statistics (MOS) approach (Glahn et al., 1972), namely to fit the statistical models using historical predictions and corresponding observations, then apply the fitted model to estimate the conditional predictive uncertainty of future observations (Hamill, 2018). Hydrological post-processing methods have two goals: 1) to estimate the conditional predictive uncertainty of the output of hydrological models with point predictions. In this sense, the post-processing methods convert deterministic predictions to probabilistic predictions in a simple way. 2) To correct the systematic bias of hydrological models and to achieve sharp hydrological predictions. We extend the hydrological post-processing concept in section 3.2.1 and 4.2.2. Here, the hydrological post-processing is applied in the context of the conditional predictive uncertainty, which characterises our best knowledge of future outcomes.

1.2 Predictive uncertainty

Predictive uncertainty describes the probability distribution of a future occurrence (predictand) given (conditional to) all the available information and knowledge that we obtain using hydrological model predictions (Biondi and Todini, 2018; Krzysztofowicz, 1999; Todini, 2008). Krzysztofowicz (1999) and Todini (2008) highlighted two central ideas. First, the goal of the forecasting is the uncertainty quantification of predictand rather than the uncertainty of predictions generated by hydrologic models. Secondly, the best way to improve forecasting is reducing predictive uncertainty. To explain these ideas and to follow Todini (2008), let us introduce the concept of a joint probability distribution of the predictand q_o and the model prediction q_s. Figure 1 illustrates the joint sample frequency of q_o and q_s that can be used to quantify the joint probability density. For any given hydrological model, the predictions q_s, will be a function of model parameters θ and of the input model forcing x (e.g. precipitation, potential evapotranspiration, etc.). Therefore, the joint probability density can be expressed as $f\left(q_o,(q_s|x,\hat{\theta})\right)$. To predict q_o , it should derive the conditional predictive distribution of q_o given q_s. It can be easily achieved cutting for a given q_s the previously mentioned joint probability density (Figure 1) and renormalising it. This can be formalised as:

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

Note that the conditional predictive uncertainty of the equation (1) indicates the predictive uncertainty of a “given” model under a “given” model forcing, “given” initial and boundary conditions and a “given” set of parameter values. So, the aforementioned conditional predictive uncertainty is not related to the uncertainty induced by the hydrological parameter values, hydrologic model choice, initial boundary conditions errors and input/output measurement errors. For such reason, in the sequel term "predictive uncertainty" refers to the "conditional predictive uncertainty". As shown in Figure 1, the conditional $f(q_o|q_s)$ is less dispersed than the marginal $f(q_o)$ given that the uncertainty will be reduced by the additional information produced by the hydrological model predictions.

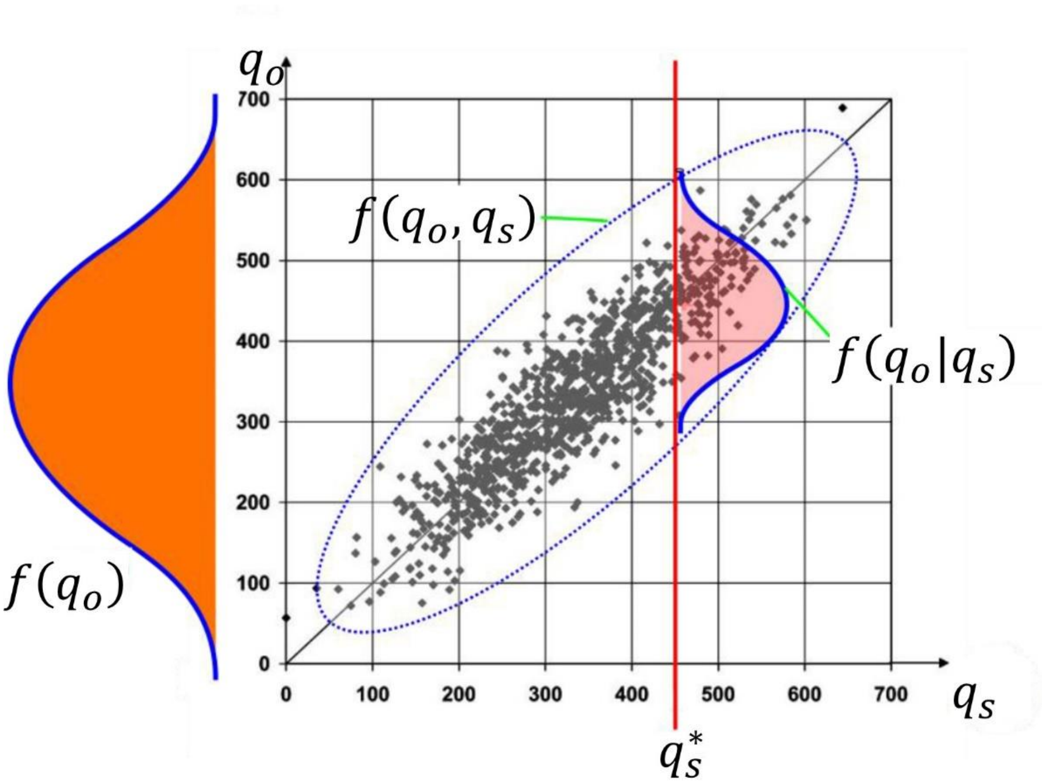

Figure 1. The predictive density defined as the probability density of the real quantity q_o conditional upon model predictions, q_s , where q_s is considered as known (namely, certain and not affected by uncertainty) at the time of prediction (Redrawn from Hernández López (2017)).

1.3 Performance metrics

There are many metrics to evaluate the performance of post-processing methods. In this thesis, we followed the verification metrics recommended by Laio and Tamea (2007), Thyer et al. (2009) and Renard et al. (2010). For deterministic metrics, the Nash-Sutcliffe Efficiency (NSE) and the Kling-Gupta Efficiency (KGE) scores were applied to assess the accuracy of the proposed post-processing methods.

1.3.1 Nash-Sutcliffe Efficiency (NSE)

NSE score measures the squared differences between the predicted q_s and observed streamflow q_o normalized by the variance of the observed flows (Equation (2)). NSE ranges between 1 (prefect fit) and $-\infty$, whereas $NSE < 0$ indicates that predictions are not superior than the observed mean as a forecast.

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

1.3.2 Kling-Gupta Efficiency (KGE)

KGE score was introduced as the modified version of the NSE by Gupta et al. (2009). KGE includes the correlation coefficient, mean bias and relative variability.

$$KGE = 1 - \sqrt{(\rho - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}\,, \tag{3}$$

where $\rho = \frac{Cov(q_o, q_s)}{\sigma_{q_o}\sigma_{q_s}}$ is the Pearson correlation, $\beta = \frac{\mu_{q_s}}{\mu_{q_o}}$ is the bias ratio and $\gamma = \frac{\sigma_{q_s}/\mu_{q_s}}{\sigma_{q_o}/\mu_{q_o}}$ is the variability ratio. All components have their optimum at unity. NSE and KGE were applied to the predictive distribution median.

For probabilistic metrics, the reliability, precision and PQQ plot were applied to evaluate the predictive uncertainty.

1.3.3 PQQ plot

The PQQ plot shows how well probabilistic forecasts represent the uncertainty in observations (Laio and Tamea, 2007; Thyer et al., 2009). In the PQQ plot context, if the predictive distribution and observed data are consistent, the corresponding p-value distribution should be uniformly distributed over the whole interval [0,1]. In other words, perfect reliable predictions are given when observed relative frequencies equal prediction probabilities, indicating in 1:1 diagonal line. Therefore, the reliability score can be derived considering the difference between the PQQ plot curve and the diagonal line.

1.3.4 Reliability

The reliability score quantifies the statistical consistency between the observed time series and the predictive distribution.

$$Reliability = \frac{2}{n} \sum_{i=1}^{n} \left| F_U - F_{q_s}(q_o) \right| \tag{4}$$

Where F_U is the uniform cumulative distribution function (CDF) and $F_{q_s}(q_o)$ is the predictive CDF. Precision metric refers to the concentration of the predictive distribution. In other terms, it refers to the spread of the predictive distribution (Renard et al., 2010).

1.3.5 Precision

Precision score is also named resolution or sharpness. The highest precision values are preferred because they indicate sharp predictive distribution.

$$Precision = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{E}[q_s]}{\sigma[q_s]} \tag{5}$$

Where $\mathrm{E}[]$ and $\sigma[]$ are the expectation and the standard deviation operators. Note that two post-processors can both yield reliable predictive distribution, but width different degrees of precision.

1.3.6 d-factor

The d-factor score indicates the average with of the prediction interval and is defined as follow:

$$d - factor = \frac{\frac{1}{n}\sum_{i=1}^{n}(Q_{up} - Q_{low})}{\sigma_{q_o}} \tag{6}$$

Where Q_{up} and Q_{low} are the upper and lower bounds of the 95% prediction interval and σ_{q_o} is the standard deviation of the observed streamflow. The d-factor near to 1 is preferred.

1.3.7 Containing ratio (CR 95%)

The containing ratio score (CR 95%) is the percentage of the observations bracketed by the 95% uncertainty band. Here, we applied the 95% prediction interval based on the 2.5 and 97.5 percentiles. Therefore, a perfect uncertainty quantification was achieved when the CR came close to 95%.

1.3.8 Average band width of 95% uncertainty band (B) and deviation amplitude (D)

The average band width of 95% uncertainty band (B) (see Equation (7)) and the average deviation amplitude (D) (see Equation (8)) assess the degree of predictions deviating from observations. D illustrates the actual discrepancy between the trajectories consisting of the points inside of the prediction bounds and the observed hydrograph. The equation of B and D are expressed as:

$$B = \frac{1}{n}\sum_{i=1}^{n}\left(Q_{up} - Q_{low}\right) \tag{7}$$

$$D = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\left|\left(Q_{up} + Q_{low}\right) - q_o\right| \tag{8}$$

1.4 Main aims and scope

This thesis is part of the research line entitled “uncertainty estimation” of the Research Group of Hydrological and Environmental Modelling (GIMHA). In general, this thesis focuses on improving hydrological post-processing methods for assessing the conditional predictive uncertainty of monthly streamflows. Although many post-processing methods have been proposed, there has not been sufficient researches that explore and compare new statistical techniques to post-process and quantify predictive uncertainty (Hamill, 2018). For instance, little literature is available on estimating the conditional predictive uncertainty with intractable likelihood, i.e. when predictions are not in synchrony with observations (Maraun, 2016) or scenarios of data scarcity. Besides, there is a clear need for assessing and comparing post-processing methods to diagnose the most suitable approaches for specific applications (Li et al., 2017).

The general objective of this thesis is to improve hydrological post-processing methods for assessing the conditional predictive uncertainty of monthly streamflows. This study advances the hydrological post-processing framework developed in Todini (2008) and further developed in Coccia and Todini (2011). This thesis deal to two issues of the hydrological post-processing scheme i) the heteroscedasticity problem, which means the prediction uncertainty increases with the magnitude of prediction variables (Coccia and Todini, 2011) and ii) the intractable likelihood problem, which means the likelihood function is unavailable in closed form or by numerical derivation (Robert, 2016). To overcome these issues, we proposed three specific objectives:

- To develop and evaluate a new post-processing method to deal with the heteroscedasticity problem (Chapter 2).
- To introduce a new post-processing method to deal with the intractable likelihood problem (Chapter 3).
- To apply the ABC post-processor to quantify the uncertainty of streamflow statistics of climate change projections (Chapter 4).

In a first step and relate to the heteroscedasticity problem, we developed and evaluated a new post-processing approach, called GMM post-processor, which is based on the Bayesian joint probability modelling approach and the Gaussian mixture models (Chapter 2). Besides, we compared the performance of the proposed post-processor with the well-known exiting post-processors for monthly streamflows across 12 MOPEX catchments (also in Chapter 2).

Secondly and related to the intractable likelihood problem, we introduced a new method to quantify the conditional predictive uncertainty in hydrological post-processing contexts when it is cumbersome to calculate the likelihood (intractable likelihood). Sometimes, it can be challenging to estimate the likelihood itself in hydrological modelling, especially working with complex models or with ungauged catchments. After a literature review, we have proposed the ABC post-processor that exchanges the requirement of calculating the likelihood function by the use of some sufficient summary statistics and synthetic datasets. Chapter 3 presents the ABC post-processor and proves its skill in two scenarios.

In a third step, we applied the ABC post-processing method to estimate the uncertainty of streamflow statistics obtained from climate change projections (Chapter 4), such as a particular case of intractable likelihood problem. Finally, Chapter 5 presents the main conclusions and provides future research lines.

CHAPTER 2. Assessing post-processing approaches for monthly streamflow in 12 MOPEX catchments

2.1 Introduction

Water resources managers demand guidance and support to make robust decisions about different water challenges, for instance, warming climate (Weaver et al., 2013), water scarcity (Mekonnen and Hoekstra, 2016; Veldkamp et al., 2017), water pollution (Allan, 2003; Vogel et al., 2015), increasing global population (Wagener et al., 2010), changing land use (Foley, 2005), water security (Vörösmarty et al., 2010; Wheater and Gober, 2015) and so forth. Researchers and engineers provide this guidance through hydrological forecasts. However, making valuable hydrologic forecasts is an awkward proceeding because of the watersheds are complex systems (Dooge, 1986), which suffer from spatial heterogeneity, scale, dynamic behaviour, and co-evolution of climate, soil, vegetation and human. In other words, the lack of comprehensive understanding of the hydrological process (*epistemic* uncertainty) and the incapacity to portray the catchment properties' heterogeneity (*aleatory* uncertainty) produce hydrological predictions with significant uncertainties (Sivapalan, 2018). These predictions involve many sources of uncertainty (Roundy et al., 2018; Vrugt and Massoud, 2018). In a broad sense, the three main sources of hydrological model uncertainty arise from data uncertainty, parameter uncertainty, and model structure uncertainty. In this context, predictive uncertainty quantification in hydrological modelling is essential to water and environmental resources risk management (McInerney et al., 2018). Thus to support decision makers, hydrological predictions need to be accompanied by an uncertainty analysis (Ehlers et al., 2019). Besides, Schoups and Vrugt (2010) stated that hydrological predictions without uncertainty analysis are impractical tools to plan and operate water resources systems. Uncertainty quantification is also considered a desirable scientific practice (Refsgaard et al., 2007). Water managers and stakeholders acknowledge that work in an environment of changing and uncertainty, so they are increasingly interesting to include the uncertainty associated with hydrological predictions in the impact of their possible decisions (Cosgrove and Loucks, 2015). The uncertainty quantification does not simplify decision-making process, but avoid it is ignore the reality (Kavetski, 2019).

In the last four decades, various uncertainty analysis methods have been developed to quantifying, reducing, and communicating the uncertainty of hydrological predictions. These methods included the generalised likelihood uncertainty estimation (GLUE), which is a simple method to estimate parametric and predictive uncertainty (Beven and Binley, 1992). GLUE has been criticised for not being formally Bayesian, i.e. the informal method makes no explicit reference to the error model (Mantovan and Todini, 2006). The Bayesian total error analysis (BATEA), which explores input uncertainties due to errors in the forcing data and multiple sources of uncertainty (Kavetski et al., 2006a, 2006b). Various methods quantify the uncertainty of the model structure (Bulygina and Gupta, 2009; Butts et al., 2004; Hoeting et al., 1999; Vrugt and Robinson, 2007). Data assimilation methods address the uncertainty of initial conditions (Moradkhani et al., 2019, 2005b). Schoups and Vrugt (2010) proposed the formal generalized likelihood function, which relaxes several statistical assumptions. Currently, the information theory approach which defines a benchmark for the best model performance (Gong et al., 2013), the approximate Bayesian computation (Fenicia et al., 2018; Kavetski et al., 2018), which uses a free likelihood and hydrological signatures and many post-processing methods (W. Li et al., 2017).

In this chapter, we focused on hydrological post-processing methods, which are well-known statistical methods that relate the observed variables (predictands) to the corresponding simulated variables (predictors) to quantify the predictive uncertainty in hydrological modelling (W. Li et al., 2017). They dilute errors from observed data, model parameters, model initial and boundary conditions, and model structure (Buizza, 2018; Ye et al., 2014). In recent years, diverse hydrological post-processing methods have been developed and reported in the scientific literature. Early, Krzysztofowicz and Kelly (2000) introduced the hydrologic uncertainty processor (HUP) to evaluate the predictive uncertainty in hydrological forecast given a set of observed and predicted time series. Montanari and Grossi (2008) proposed a meta-Gaussian based on multivariate regression to match streamflow errors to predictors variables. Because all models are imperfect and no single model is the best under all circumstances, Raftery et al. (2005) established the Bayesian model average (BMA), which is a multi-model method. Based on the Bayesian paradigm, Todini (2008) presented the model conditional processor (MCP), and Wang et al. (2009) developed the Bayesian joint probability (BJP) model for seasonal streamflow forecasting. In the same line, Zhao et al. (2011) introduced the general linear model post-processor (GLMPP). Weerts et al. (2011) proposed the quantile regression approach to avoid any assumptions in the regression. Currently, many methods were developed, e.g., a post-processing with error model (Evin et al., 2014; Woldemeskel et al., 2018), non-parametric post-processing methods (Brown and

Seo, 2013), data-driven resampling techniques (Ehlers et al., 2019; Sikorska et al., 2015; Solomatine and Shrestha, 2009), copula post-processing methods (Klein et al., 2016; Madadgar and Moradkhani, 2014; Schefzik et al., 2013). This list of post-processing methods is not meant to be exhaustive but for a detailed review, see Li et al. (W. Li et al., 2017). The errors of hydrological modelling generally are autocorrelated, non-normal and heteroscedastic (McInerney et al., 2017; Schoups and Vrugt, 2010; Smith et al., 2015). To address this issue, many of the post-processors apply transformation methods, e.g., the Normal Quantile Transformation (NQT) (Bogner et al., 2012), the Box-Cox transformation (Box and Cox, 1964), the log–sinh transformation (Wang et al., 2012) and so forth.

There is a clear need for assessing and comparing post-processing methods to diagnose the most suitable approaches for specific applications (W. Li et al., 2017). As stated during a 2016 workshop on Statistical post-processing, recommendation # 2: “*Performance more comparisons of existing algorithms to determine which are the most skilful and reliable*” (Hamill, 2018). Several studies have compared different post-processing methods for hydrological variables. For instance, Van Andel et al. (2013) reported on the intercomparison experiment for post-processing techniques that has been initiated in 2011 by the International Community on Hydrologic Ensemble Predictions indicating preliminary that post-processing methods revealed different behaviour. Schepen and Wang (2015) compared the performance of the Bayesian model averaging (BMA) and quantile model averaging (QMA) to merge statistical and dynamic forecasts for seasonal streamflows in 12 Australian catchments finding that both methods performed similarly. Klein et al. (2016) compared the predictive skill of the copula uncertainty processor (COP) based on pair-copula construction, Bayesian model averaging (BMA), quantile regression and model conditional processor (MCP) using the multivariate truncated normal distribution for daily streamflows recommending the COP. Recently, Woldemeskel et al. (2018) evaluated post-processing approaches using three transformations, namely logarithmic, log-sinh and Box-Cox over 300 Australian catchments for monthly and seasonal streamflow forecasts, concluding that the Box-Cox transformation with $\lambda = 2$ was the best-performing post-processing method, especially in dry catchments. Muhammad et al. (2018) tested four statistical post-processing techniques: linear regression (LR), quantile mapping (QM), QMA, and BMA for seasonal streamflows, establishing that no post-processor outperformed other methods. Finally, Sharma et al. (2018) compared the relative effects of statistical pre-processing and post-processing on a regional hydrological ensemble prediction system.

In this research, we propose a new statistical post-processing method that combines the Bayesian joint probability (BJP) modelling approach and the Gaussian mixture models (GMM) to quantify the conditional predictive uncertainty in monthly streamflows. We also compare the performance of the proposed post-processor, which is called GMM post-processor, to existing Bayesian post-processors including the Model Conditional Processor (MCP), the MCP using the truncated Normal (MCPt) and the linear regression post-processor using Markov Chain Monte Carlo (MCMC) to inference parameters across 12 MOPEX catchments. However, the use of Gaussian mixture is not new in the hydrological post-processing community. Recently, Feng et al. (2019) and Klein et al. (2016) used Gaussian mixture for estimating the marginal distribution of hydrological post-processing approaches. On the other hand, the proposed GMM post-processing method merges the BJP and GMM to model the joint probability distribution that describes the relationship between deterministic hydrological predictions (predictors) and corresponding observed streamflows (predictands). This approach has never been used before in this situation, and it is interesting to test the merits of GMM post-processing method.

2.2 Methodology

2.2.1 Review of hydrological post-processing methods

A hydrological post-processor is a statistical technique used to improve typically deterministic forecasts by relating hydrological model outputs to observations (Ye et al., 2014). Since the hydrological modelling process has many sources of uncertainty, post-processing is required to characterise these uncertainties and remove systemic bias in the predicting process (Hopson et al., 2019). Here, the hydrological post-processing was applied in the context of the predictive uncertainty, which characterises our best knowledge of future outcomes. Mainly, the predictive uncertainty describes the probability of predictand (streamflow, water level, soil moisture, etc.) conditional over all the information that we obtain using hydrological models (Krzysztofowicz and Kelly, 2000) (see Section 1.2). In this chapter, we compared the performance of the proposed GMM post-processing method to well-known Bayesian post-processors as the traditional model conditional processor (MCP), MCP using the truncated Normal (MCPt) for dealing with the heteroscedasticity and the linear regression post-processor using MCMC to inference parameters.

2.2.1.1 Model Conditional Processor (MCP)

Todini (2008) proposed the model conditional processor (MCP), which is a Bayesian joint probability model (BJP) approach developed for estimating the real-time flood predictive uncertainty. MCP can be applied for a univariate model approach, multivariate model approach, multiple lead-time, and multi-temporal approach (Coccia and Todini, 2011). MCP has been used in a meteorological reanalysis ensemble (Reggiani et al., 2016), a multi-temporal approach for real-time forecasting (Barbetta et al., 2016), hydro-meteorological ensembles (Biondi and Todini, 2018), and for complement satellite rainfall information (Massari et al., 2019). Notably, MCP established a joint probability distribution to describe the relationship between deterministic hydrological predictions (predictors) and corresponding observed streamflows (predictands). The joint distribution was modelled as a bivariate normal distribution after transformation of the marginal distributions using a non-parametric distribution. The MCP implementation was based on three main steps. First, the data transformation was handled using the Normal Quantile Transformation (NQT) (Bogner et al., 2012), which is a non-parametric technique to transform the predictions and observations into the Gaussian or normal space. Second, the conditional predictive distribution was computed using the Bayes formula in the context of the Bayesian inversion, in which the predictions and observations are available at the same time. If a predictand q_o was transformed to η_o and a predictor q_s was transformed to η_s, the relationship between η_o and η_s was formulated by a bivariate normal distribution:

$$\begin{bmatrix}\eta_o \\ \eta_s\end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{9}$$

where $\boldsymbol{\mu} = \begin{bmatrix}\mu_{\eta_o} \\ \mu_{\eta_s}\end{bmatrix}$ is the vector of means 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. Third, the predictive uncertainty in the normal space was finally reconverted to the real space through the inverse NQT. As in hydrological modelling process, we applied a split sample approach. During the calibration period, the joint and marginal distribution was identified for Bayes theorem application while during the validation period, the MCP model was conditioned on new predictor values. For a particular parameter set, θ, and new transformed predictor value η_{s_new},

$$\eta_{o\_new}|\eta_{s_{new}},\theta \sim N\left[\mu_{\eta_o}+\rho_{\eta_o\eta_s}\frac{\sigma_{\eta_o}}{\sigma_{\eta_s}}\left(\eta_{s_{new}}-\mu_{\eta_s}\right),\sigma_{\eta_o}^2\left(1-\rho_{\eta_o\eta_s}^2\right)\right] \tag{10}$$

Interestingly, the MCP approach was simpler to implement and had a lower computational cost because it applied the analytical treatment of the bivariate normal distribution. In this chapter, MCP required as input monthly streamflow observations and the corresponding hydrological predictions from the GR4J model. In addition, MCP was used in a univariate framework, i.e., just one predictor (hydrological model's output) to estimate the conditional predictive uncertainty of monthly streamflow observations. For more details on MCP, we refer the readers to Todini (2008).

2.2.1.2 MCP using the truncated Normal distribution (MCPt)

The traditional MCP approach assumes that the error variance is homoscedastic (constant variance). However, this assumption is hard to justify for many hydrological variables, for example, low and high streamflows can show higher variance than their mid-range. To address this assumption, Coccia and Todini (2011) extended the traditional MCP splitting the entire Normal domain into two or more subdomains where Truncated Normal Distribution (TND) can be applied. Therefore, the joint distribution in the Normal space was not unique so that it could split into two or more (TND). Coccia and Todini (2011) recommended two TNDs to describe the heteroscedasticity of variance adequately. Although initially the MCPt with truncated Normal was developed for the real-time flood, we applied the MCPt approach to estimate the conditional predictive uncertainty of monthly streamflows.

2.2.1.3 GMM post-processor

The GMM post-processing method is a Bayesian joint probability (BJP) modelling approach and shares the same structure and philosophy of the traditional MCP from section 2.2.1.1, but GMM post-processor proposes a diverse method to address the heteroscedasticity of variance, where the error variance is characterised using Gaussian mixture models. As GMM is a class of data-driven model, it provides great flexibility in fitting probability models with complex characteristics, but the price for this flexibility is an increase in the number of parameters. GMM provides a semiparametric approach to model the unknown probability model represented as a weighted sum of Gaussian components (McLachlan and Peel, 2000). In particular, GMM has much of the flexibility of non-parametric approaches, while retaining some of the advantages of parametric approaches, such as keeping the dimension of the parameter space down to a practical size (Melnykov and Maitra, 2010). Let $\boldsymbol{Y} = (\boldsymbol{Y}_1^T, \cdots, \boldsymbol{Y}_n^T)^T$, denote a random sample of size n, where $\boldsymbol{Y_j}$ is a p-dimensional random vector with probability density function $f(\boldsymbol{y}_j)$ on $\mathbb{R}^p$. Note that $\boldsymbol{Y}$ represents the entire sample, where a realisation of an observed random sample is denoted by $\boldsymbol{y} = (\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_n^T)^T$. The density $f(\boldsymbol{y}_j)$ of $\boldsymbol{Y_j}$ can be written in the form

$$f(\boldsymbol{y}_j) = \sum_{i=1}^{k} \pi_i \, f_i(\boldsymbol{y}_j|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{i=1}^{k} \pi_i \, N_i(\boldsymbol{y}_j|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\,, \tag{11}$$

where π_i are called the mixing proportions or weights, are non-negative quantities that sum to one, the $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean vector and covariance matrix of the *k-th* Gaussian component, respectively. The distribution of each individual Gaussian component is

$$f(\boldsymbol{y}_j|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{exp\left[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu}_k)^T \cdot \boldsymbol{\Sigma}_k^{-1} \cdot (\boldsymbol{y} - \boldsymbol{\mu}_k)\right]}{(2\pi)^{p/2}\sqrt{|\boldsymbol{\Sigma}_k|}} \tag{12}$$

We applied the Expectation-Maximization (EM) algorithm (McLachlan and Krishnan, 2008), which assumes a priori number of Gaussian components, to identify the parameters of GMM, i.e. π_i, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$. Usually, for monthly streamflow variability, three Gaussian components are recommended. Note the GMM has heteroscedastic components (components with unequal variance). Given the great flexibility of GMM, we applied the Gaussian Mixture model for modelling of heterogeneity in monthly streamflows. In the GMM approach, each component of a Gaussian mixture density is associated with a group or cluster, i.e. low, medium and high streamflows. We also implemented a hard clustering, which is a technique that assigns each data point to exactly one cluster. For GMM post-processor, cluster assigns each point to one of the three mixture components in the Gaussian Mixture Model. The centre of each cluster is the matching mixture component mean. In forecasting mode, it is necessary to remember that GMM post-processor has no extrapolation ability as a data-driven model. For a particular parameter set, θ, and new predictor follow the equation 10.

2.2.1.4 Linear regression post-processor

Linear regression is a common statistical approach to correct forecast biases and to provide predictive uncertainty in deterministic predictions (Siegert and Stephenson, 2019). Linear regression is one of the earliest statistical post-processing procedure, and it is named Model Output Statistics (MOS) in climatology literature. In predictive uncertainty, we used the linear regression to establish a statistical correlation between the predictand (observations) and the predictors (model outputs). Especially, linear regression post-processor predicts future observations (also their predictive uncertainty) with future hydrological predictions. The linear regression post-processor is as follows:

$$\boldsymbol{y}^o = \beta_0 + \beta_1 \boldsymbol{y}^s + \boldsymbol{\varepsilon}, \tag{13}$$

where vectors $\boldsymbol{y}^o$ and $\boldsymbol{y}^s$ are the observed streamflows (predictands) and streamflow predictions (predictors), respectively; β_0 and β_1 are unknown post-processor parameters, and $\boldsymbol{\varepsilon}$ is an error term to account for random sampling noise. It is assumed that $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$ is identical independent distributed (i.i.d). The linear regression post-processor uses an aggregational approach to describe errors, and all the sources of uncertainty are grouped into a linear residual error equation (13). Three components on the right-hand side denote three distinct sources of estimation errors. The first term, β_0, indicates constant deviation and can be called the displacement error. The second term, with an error parameter β_1, represents a scale error or dynamic-range error. The third term is a random error $\boldsymbol{\varepsilon}$, which symbolises aleatory uncertainty, and the first two terms characterise the systematic or epistemic error. We applied the Normal Quantile Transformation (NQT) (Van Der Waerden, 1953) to satisfy the (i.i.d) assumptions. To estimate the uncertainty, we used Bayesian modelling, so the posterior distribution of the statistical post-processor equation (13) is, $(\boldsymbol{\theta} | \boldsymbol{y}^o, \boldsymbol{y}^s)$, which is given by Bayes Theorem:

$$p(\boldsymbol{\theta} | \boldsymbol{y}^o, \boldsymbol{y}^s) = \frac{p(\boldsymbol{y}^o | \boldsymbol{\theta}, \boldsymbol{y}^s) p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}^o | \boldsymbol{\theta}, \boldsymbol{y}^s) p(\boldsymbol{\theta}) \, d\boldsymbol{\theta}}, \tag{14}$$

where $p(\boldsymbol{\theta})$ is a prior parameter distribution and $p(\boldsymbol{y}^o | \boldsymbol{\theta}, \boldsymbol{y}^s)$ is a likelihood function. The linear regression post-processor assumes flat uniform priors for $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2)$ and from the assumptions in the model (13), it follows that $\boldsymbol{Y}^o | \boldsymbol{\theta}, \boldsymbol{y}^s \sim N(\mu = \beta_0 + \beta_1 \boldsymbol{y}^s, \ \sigma^2)$.

After defined the linear regression post-processor in the Bayesian framework, we can proceed to statistical inference to get parameter values and their uncertainty. In Bayesian statistics, model parameters are random variables, so we get the posterior distribution of parameters instead of one value. Usually, equation (14) does not have an analytical solution. For this reason, numerical methods based on Monte Carlo simulation are needed. For the numerical implementation of Bayesian inference, we are interested in generating a sample from the posterior distribution (see equation (14)), and properties of the distribution are approximated by properties of the sample (Kattwinkel and Reichert, 2017). Here, we applied an adaptive Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm (Haario et al., 2001), which is a numerical method for sampling any probability distribution (Kavetski, 2019), to perform the Bayesian inference. This algorithm has been shown to perform adequately in hydrologic problems (Smith and Marshall, 2008). The MCMC algorithm was implemented in R (Core Team, 2013) using the package "MHadaptive" (Chivers, 2012). Because we are more interested in the predictive uncertainty, let $\widetilde{y^o}$ be a future observation for linear regression post-processor (see equation (13)), then the posterior predictive density of a future observation $p\left(\widetilde{y^o}|\boldsymbol{y}^s\right)$, is given by:

$$p\left(\widetilde{y^o}|\boldsymbol{y}^s\right) = \int_{\emptyset} p\left(\widetilde{y^o}|\boldsymbol{\theta}\right)p(\boldsymbol{\theta}|\boldsymbol{y}^o, \boldsymbol{y}^s)\, d\boldsymbol{\theta} \tag{15}$$

Equation (15) represents both the uncertainty of the post-processor and the uncertainty due to variability in future observations (Yoon et al., 2010). In this study, the adaptive Metropolis-Hastings algorithm was implemented and ran until the convergence of the parameter posterior distribution (statistical model) was achieved. Convergence was determined by both the visual trace plot evaluation of the posterior chains and the Gelman and Rubin (1992) R statistic, which considers convergence in terms of the variance with a single chain and the variance between multiple parallel chains. Finally, we computed the conditional predictive distribution.

2.2.2 MOPEX database

Data used in this research was the observed and predicted daily streamflow from the Second Workshop on Model Parameter Estimation Experiment (MOPEX) database (Duan et al., 2006; Ye et al., 2014). In the MOPEX database, we selected twelve catchments spread over the South-eastern quadrant of the

United States for application and test (Figure 2). The aridity ratio ranged from 0.43 to 2.22 (Table 1), and the run-off ratio ranged from 0.15 to 0.63; therefore, 12 catchments represented different hydroclimatic conditions (Figure 3). Catchment information is shown in Table 1, and catchment locations are mapped in Figure 2.

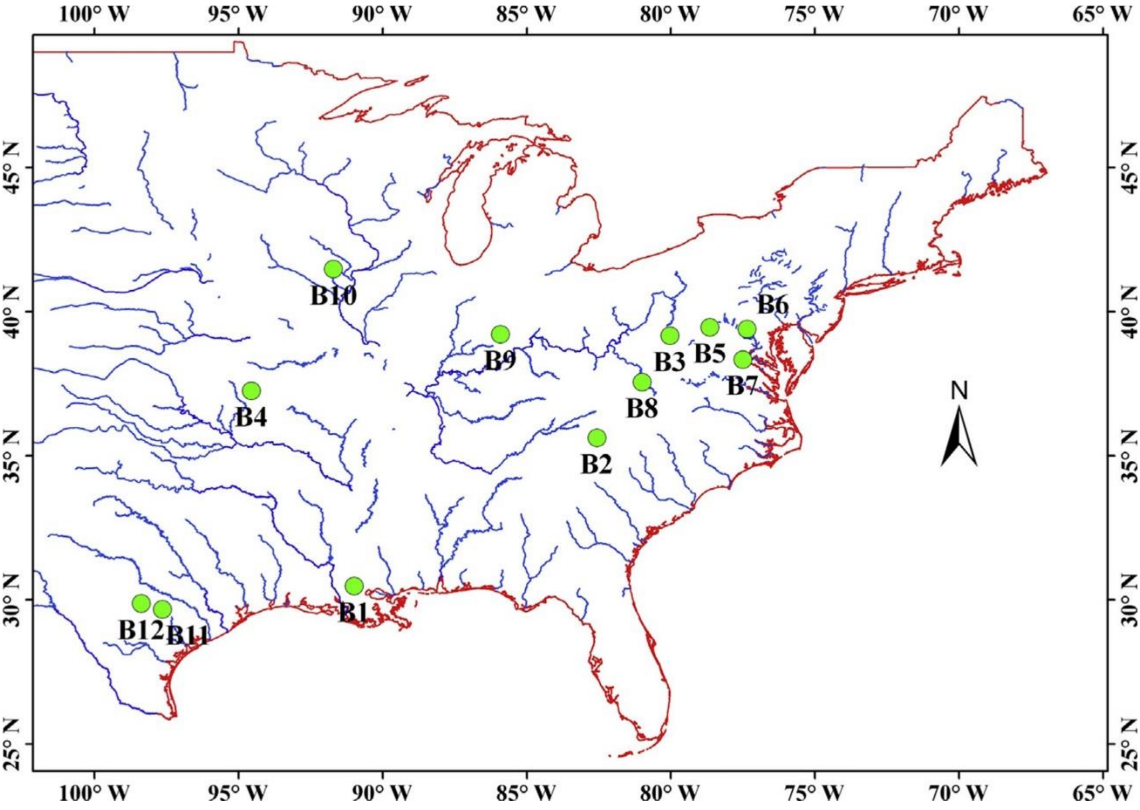

Figure 2. Location of the 12 MOPEX catchments. Take from Ye et al. (2014).

Table 1. Hydrological information of the 12 MOPEX catchments.

ID	Station name	Elev.	Area (km^2)	P	PET	Q	Run-off ratio (Q/P)	Aridity ratio (PET/P)
B1	Amite River Near Denham Springs, LA	0	3315	1560	1068.49	612	0.39	0.67
B2	French Broad River At Asheville, NC	594	2448	1378	588.89	795	0.58	0.43
B3	Tygart Valley River At Philippi, WV	390	2372	1164	661.36	736	0.63	0.57
B4	Spring River Near Waco, MO	254	3015	1075	1119.79	300	0.28	1.04
B5	S Branch Potomac River Nr Springfield, WV	171	3810	1043	635.98	339	0.33	0.61
B6	Monocacy R At Jug Bridge Nr Frederick, MD	71	2116	1042	906.09	421	0.4	0.87
B7	Rappahannock River Nr Fredericksburg, VA	17	4134	1028	856.67	375	0.36	0.83
B8	Bluestone River Nr Pipestem, WV	465	1020	1017	678.00	419	0.41	0.67
B9	East Fork White River At Columbus, IN	184	4421	1014	838.02	377	0.37	0.83
B10	English River At Kalona, IA	193	1484	881	989.89	261	0.3	1.12
B11	San Marcos River At Luling, TX	98	2170	819	1462.50	170	0.21	1.79
B12	Guadalupe River Nr Spring Branch, TX	289	3406	761	1691.11	116	0.15	2.22

Elev: elevation (m), P: mean areal precipitation (mm/year), PET: potential evapotranspiration (mm/year), Q: observed streamflows (mm/year).

Figure 3 shows the fluctuating control of available water and accessible energy on the partitioning of precipitation between evaporation and run-off for each of 12 MOPEX catchments. Please note that the evaporation in the French Broad catchment is constrained by the annual supply energy (B2 in Figure 3), while in the Guadalupe catchment is constrained by the annual supply water (B12 in Figure 3).

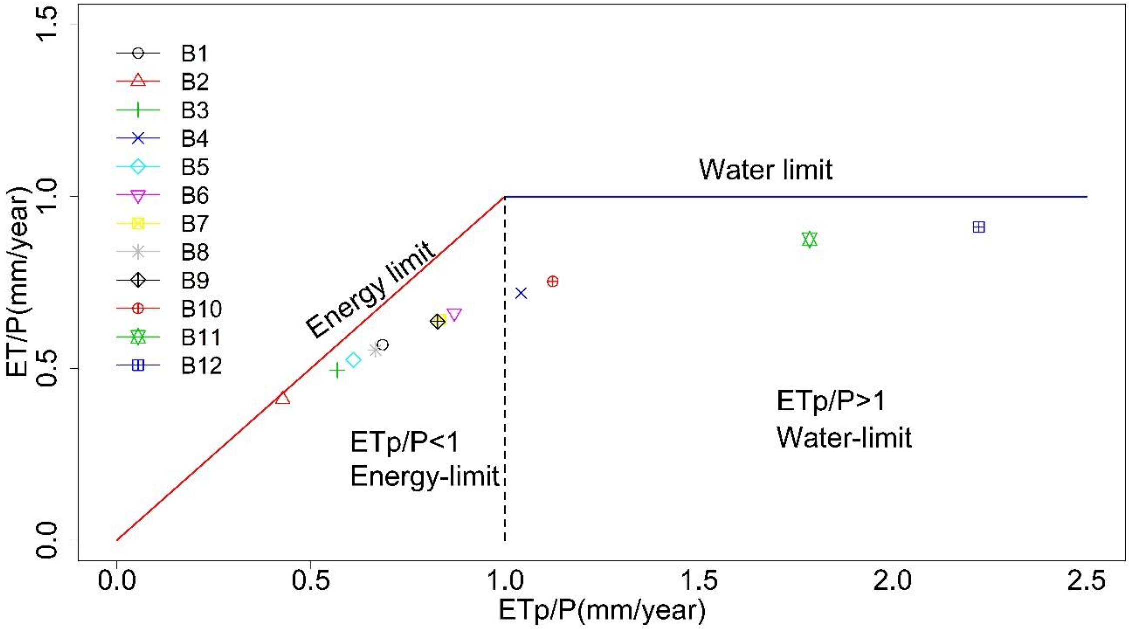

Figure 3. Budyko curve for the 12 MOPEX catchments.

2.2.3 GR4J rainfall-runoff model

There were seven different hydrological models in the MOPEX database, but we just selected the GR4J model predictions because of we were interested in the performance of the post-processing methods instead of hydrological models, and GR4J is a well-known model and widely used for the hydrological modelling. The GR4J rainfall-runoff model is an efficient and parsimonious daily lumped hydrological model described by Perrin et al. (2003).

2.2.4 Verification metrics

There are many sets of verification metrics that could be adopted for post-processor assessment. Because we focused on the predictive uncertainty quantification, we applied deterministic and probabilistic measures, which examine the accuracy, reliability and robustness. We followed the verification metrics recommended by (Laio and Tamea, 2007; Renard et al., 2010; Thyer et al., 2009). As for deterministic measure, we used the Nash-Sutcliffe efficiency (NSE) applying to the predictive distribution median. As for probabilistic measures, we used the PQQ plot, precision, containing ratio (CR 95%) and d-factor. A predictive uncertainty can be considered accurate if it contains all the observations within the uncertainty bands. However, if the uncertainty bands are so large that there is little precision in the predictive uncertainty, the predictive uncertainty is useless for any meaningful decision-making application (Franz and Hogue, 2011). Since these verification metrics have been described in the section 1.3, we just mentioned them. For additional information, we refer to Franz and Houge (2011), Laio and Tamea (2007) and Renard et al. (2010).

2.2.5 Comparison framework

To post-process monthly streamflow predictions and to quantify their conditional predictive uncertainty, we ran the daily streamflow predictions from the GR4J model, which was previously calibrated and validated by Ye et al. (2014) for the 12 MOPEX catchments. In other words, we applied the outputs from the GR4J model as inputs for the hydrological post-processing methods. Then, we aggregated these daily predictions to monthly data because the application of post-processing methods was for water resources management. These monthly streamflow predictions were called “uncorrected” streamflow predictions (deterministic predictions). Also, we applied a split sample approach for calibrating and validating post-processor parameters using twenty years (1960-1980) for the calibration period and seventeen years (1981-1998) for the validation period. All evaluated post-processing methods applied the NQT transformation with non-parametric distribution to estimate the marginal distribution of the random variables and to move to normal space. The four post-processors were applied to the 12 MOPEX catchments separately to obtain the best post-processing methods. We chose 12 MOPEX catchments because they were the same catchments used in many hydrological studies, i.e., (Clark et al., 2008; Franz and Hogue, 2011; Kavetski and Clark, 2010; Ye et al., 2014). Assessing across a variety of different hydrologic environments helps to arrive at conclusions that are more general and to evaluate the usefulness of the post-processing methods (Gupta et al., 2014).

2.3 Results and discussion

This section outlines the results of post-processing methods evaluated in this study. We focus the results in the validation period because it is more critical for meaningful applications. In the remainder of the study, the term “uncorrected” refers to streamflow predictions obtained using the GR4J model (deterministic prediction), and the term “post-processed” refers to predictions based on a streamflow post-processing method, which includes the traditional Model Conditional Processor (MCP) from Section 2.2.1.1, MCP using the truncated Normal (MCPt) from Section 2.2.1.2, GMM post-processor from Section 2.2.1.3 and the linear regression post-processor using MCMC to inference parameters (MCMC) from Section 2.2.1.4. First, we evaluated the performance of uncorrected and post-processed streamflow predictions. Second, we presented a comparison of post-processor methods. As a general view of seasonal variation in climatic conditions over the diverse catchments, Figure 4 illustrates the long-term mean monthly uncorrected and post-processed streamflow predictions. Note that the uncorrected and post-processed streamflow predictions mimic observed data quite well. This finding evidences the high quality of the GR4J model outputs (black line).

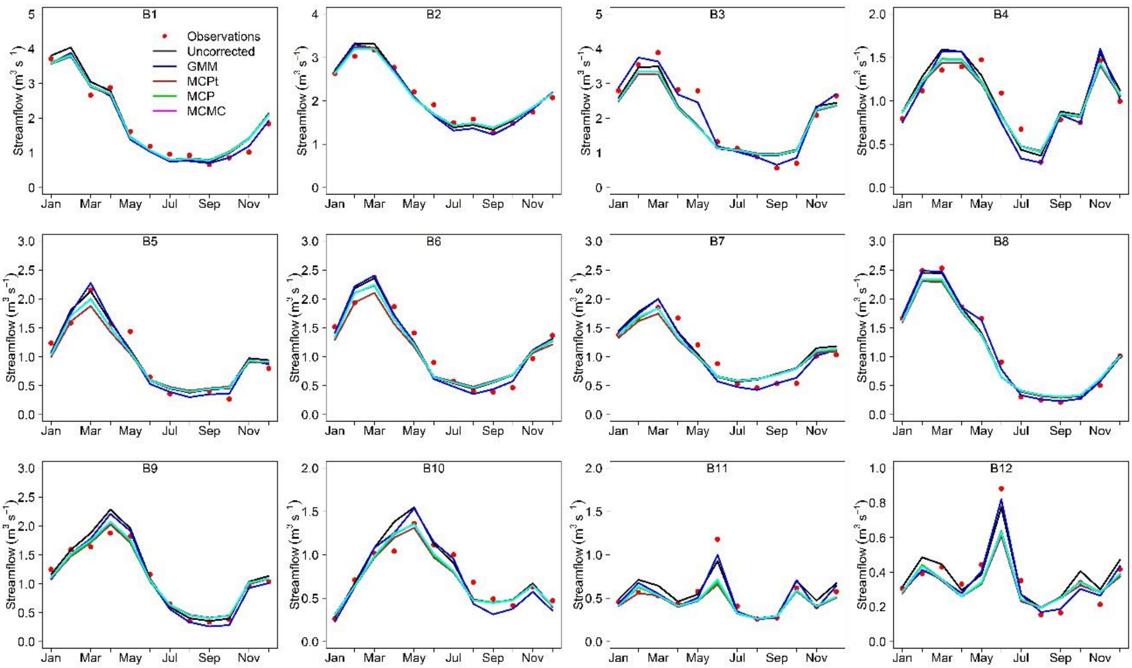

Figure 4. Median hydrographs during the validation period (1980-1998) for the 12 MOPEX catchments by observations, uncorrected predictions and four post-processors. MCP: model conditional processor, MCPt: MCP using the truncated Normal, GMM: GMM post-processor, and MCMC: linear regression post-processor using MCMC to inference parameters.

2.3.1 Comparison of uncorrected and post-processed streamflow predictions

Because uncorrected streamflow predictions were deterministic predictions, we should compare to the predictive distribution median of post-processed predictions. We applied the Nash-Sutcliffe efficiency (NSE) because it is a familiar verification metric for the hydrological modelling community. Next, we computed the percent changes in the NSE to compare the performances. The percent changes in the NSE obtained for each post-processor were compared with uncorrected predictions in the validation period (1980-1998) to evaluate the effectiveness of each post-processor. The percent change was calculated as:

$$percent\ change(\%) = \frac{NSE_{post} - NSE_{unc}}{NSE_{unc}} \times 100\ ,$$

where NSE_{unc} and NSE_{post} are the NSE of uncorrected and post-processed streamflow predictions, respectively. Figure 5 displays the percent changes in NSE computed for four post-processors during the validation period at the basins studied. In general, we see small improvements as a result of the post-processing. This relative reduced performance of post-processing can be attributed to the satisfactory predictions of the GR4J model. This finding is consistent with previous results obtained by Romero-Cuellar et al. (2019), who found that NSE increases by 25.84% post-processed poor predictions and 1.8% post-processed good predictions. Besides Ye et al. (2014), Bogner et al. (2016) and Woldemeskel et al. (2018) confirmed that post-processing methods could improve forecast significantly when uncorrected predictions are exceptionally poor. Therefore, the quality of uncorrected predictions affected the performance of post-processing because they are the inputs of post-processing methods. Figure 5 also illustrates the most considerable negative percent changes (NSE), especially in dry catchments (B4, B10, B11 and B12). This result is in line with Ye et al. (2014), who did not recommend post-processing methods for dry basins. Sometimes, post-processing methods did not improve the quality of predictions, but they provided the predictive uncertainty, which is useful and valuable information for support decision-making. Also, using only deterministic verification metrics removed a significant amount of predictive uncertainty information from the evaluation process and can lead to incorrect conclusions (Franz and Hogue, 2011), so we applied probabilistic metrics in the following sections. Interestingly, GMM post-processor produces performance improvements for all catchments, where the percent changes in the NSE ranged from 0.21 (B1) to 14.87 (B3). This finding falls in line with the results of Woldemeskel et al. (2018), who found that the post-processing scheme using Box-Cox transformation with $\lambda = 0.2$ provided improvements for monthly and seasonal streamflow forecasts across 300 Australian catchments.

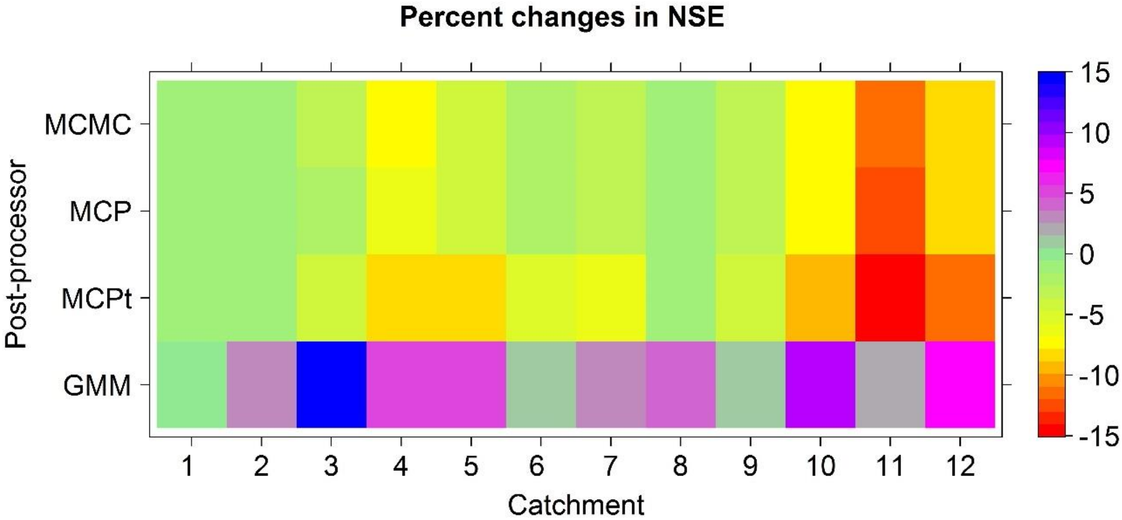

Figure 5. Percent changes in NSE computed for four post-processors during the validation period (1980 - 1998) for 12 MOPEX catchments. MCP: model conditional processor, MCPt: MCP using the truncated Normal, GMM: GMM post-processor, and MCMC: linear regression post-processor using MCMC to inference parameters.

2.3.2 Comparison of the predictive performance of post-processing methods

Figure 6 illustrates the comparison of deterministic and probabilistic metrics computed for four post-processors during the validation period (1980-1998) for 12 MOPEX catchments. Note that in Figure 6, the horizontal axis indicates different catchments, while the vertical axis indicates different post-processors. The next sections explains the performance of hydrological post-processors in terms of accuracy, reliability and precision.

2.3.2.1 Accuracy

In this chapter, we chose the Nash-Sutcliffe efficiency (NSE) to assess the accuracy skill of the post-processing methods (predictive distribution median). NSE values range from negative infinity to one (perfect skill). In terms of NSE, the performance of post-processors are considered good (NSE > 0.75) according to the range proposed by Martinez and Gupta (2010). An exciting result from Figure 6 (top left) is that the lowest skills are for dry catchments (B4, B10, B11 and B12), except in B5, which is a wet catchment. This relatively poor performance in dry catchments can be attributed to GR4J model predictions, which is not adequate for dry catchments (model structure). Clark et al. (2008) remarked that the choice of the model structure is crucial for realistic hydrological predictions. The performance of MCP, MCPt and MCMC are generally quite similar, but GMM post-processor shows the best performance in all catchments (Figure 7). This finding suggests that the Gaussian mixture models approach resolving the heteroscedasticity of monthly streamflows, especially, in dry catchments.

Conversely, Ye et al. (2014) found that the post-processing cannot improve hydrological forecast and reduce uncertainty in dry catchments because they have many zeros values in the data. Note that in contrast to the findings from Ye et al. (2014), our result suggests that the Gaussian mixture models can overcome the problem of the dry catchments. One possible explanation is that GMM provides excellent flexibility for estimating the joint probability distribution, and each component of a Gaussian mixture density is associated with a group or cluster, i.e. low, medium and high streamflows.

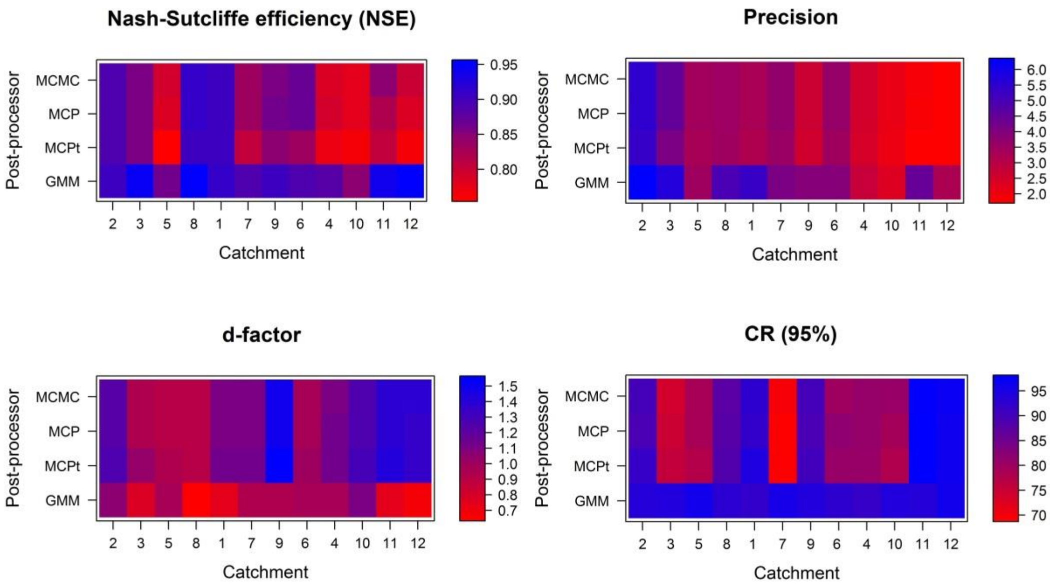

Figure 6. Comparison of deterministic and probabilistic metrics computed for four post-processors during the validation period (1980 - 1998) for 12 MOPEX catchments. MCP: model conditional processor, MCPt: MCP using the truncated Normal, GMM: GMM post-processor, and MCMC: linear regression post-processor using MCMC to inference parameters. Catchments are ordered from wet to dry regimens.

2.3.2.2 Precision

Precision metric refers to the concentration of the predictive distribution. In other words, it refers to the spread of the predictive distribution (Renard et al., 2010). Precision is also named resolution or sharpness. The highest precision values are preferred because this means sharp predictive distribution. In terms of precision, again, the most salient feature of Figure 6 (top right) is that post-processing performance is generally reduced in dry catchments (B4, B10, B11 and B12) than in wet catchments. Besides, in almost all catchments, GMM is usually sharper than other post-processors (Figure 7, top right).

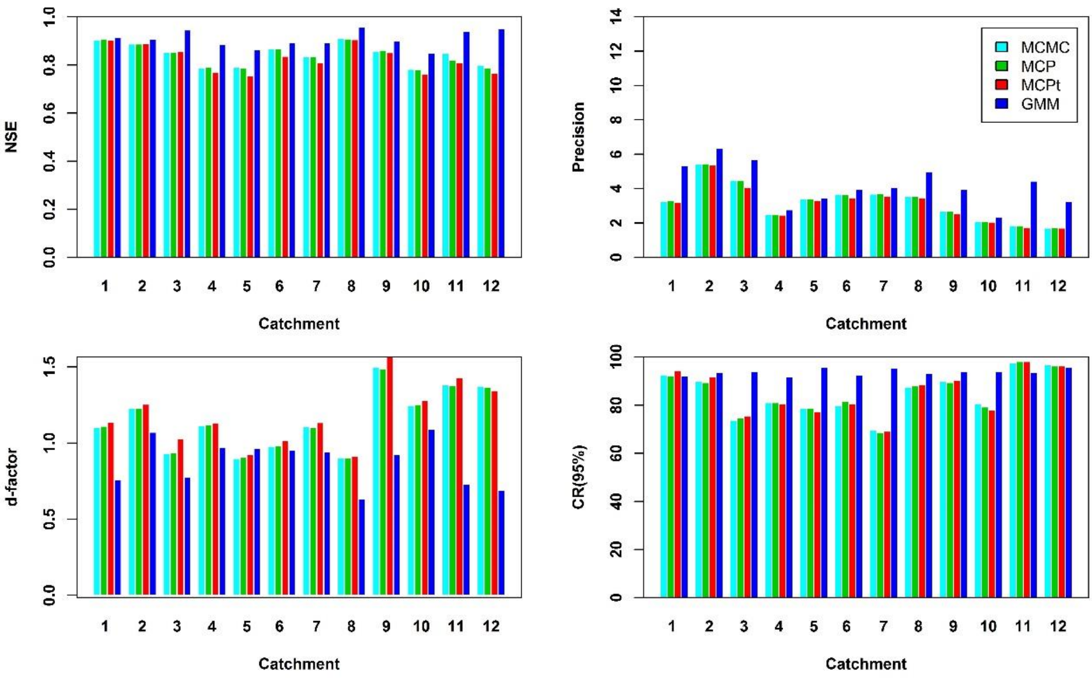

Figure 7. Comparison of deterministic and probabilistic metrics computed for four post-processors during the validation period (1980 - 1998) for 12 MOPEX catchments.

2.3.2.3 Containing ratio (CR 95%) and d-factor

The containing ratio (CR 95%) is the percentage of the observations bracketed by the 95% uncertainty band. Here, we applied the 95% prediction interval based on the 2.5 and 97.5 percentiles. Therefore, a perfect uncertainty quantification was achieved when the CR came close to 95%. In the same line, the d-factor indicates the average width of the prediction interval. The d-factor values close to 1 are preferred (Sun et al., 2017). The d-factor is another measurement of sharpness. Both the CR (95%) and d-factor reveal that the GMM is superior to all other post-processor (Figure 6 and Figure 7). This result is in coherency with previous verifications metrics, but we do not identify the reduced performance for dry basins. We analyse this finding in the next section because the reliability and PQQ-plots reveal whether the predictive distribution is over- or underestimating the observations.

2.3.2.4 Reliability

Reliability quantifies the statistical consistency between the observed time series and the predictive distribution. To assess the reliability of post-processors, we used the PQQ-plot, which was recommended by (Laio and Tamea, 2007; Renard et al., 2010; Thyer et al., 2009). In the PQQ plot context, if the predictive distribution and observed data are consistent, the corresponding p-value distribution should be uniformly distributed over the whole interval [0,1]. In other words, perfect reliable predictions are given when observed relative frequencies equal prediction probabilities, indicating in 1:1 diagonal line. Figure 8 presents the PQQ plots for four post-processors across all MOPEX basins. It is clear from this figure that the GMM curves closely follow the bisector lines for all catchments. This finding suggests that the predictive uncertainty of post-processed streamflows are reliable (blue line). In terms of reliability, the predictive uncertainty of MCP, MCPt and MCMC are very similar across all basins (Figure 8, red, green and magenta lines). Indeed, post-processed predictions produces correct uncertainty estimation, except in B5 and B11, whose uncertainty is underestimated and overestimated, respectively. These findings suggest that although post-processed predictions are ineffective in correcting biases, especially in dry basins (Figure 5), they produce a correct estimation of the predictive uncertainty (Figure 8).

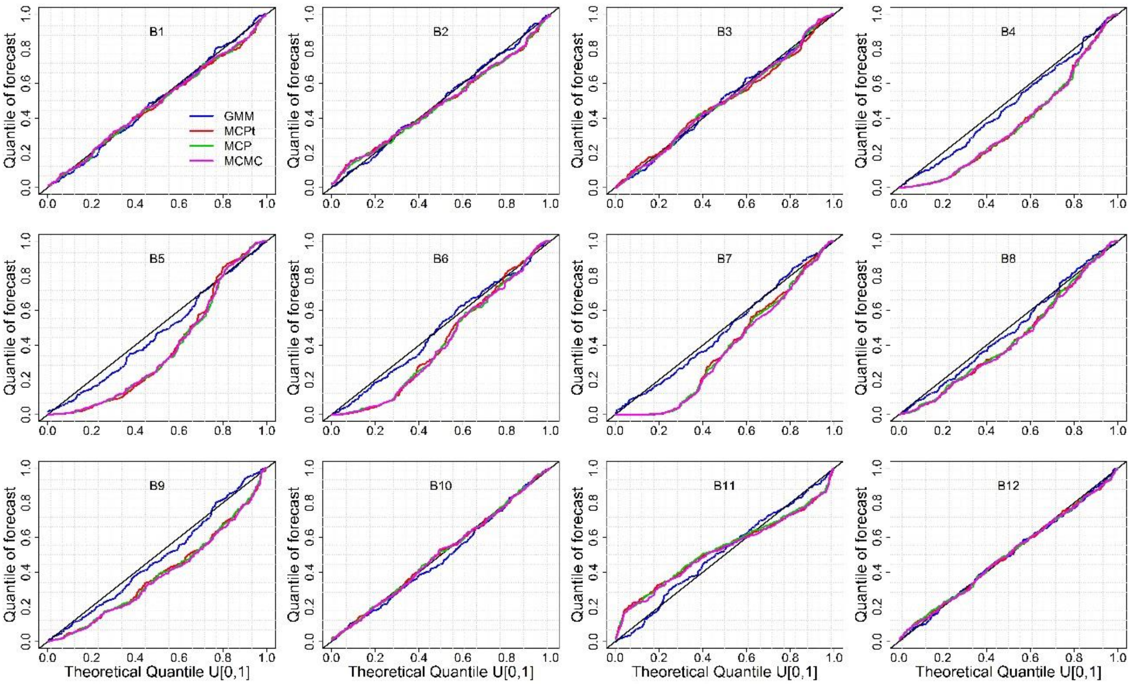

Figure 8. Conditional predictive PQQ-plot during the validation period (1980-1998) for the 12 MOPEX catchments by four post-processors.

We also examine the relationship between the performance of the post-processing methods and the hydro-meteorological processes in the catchments, but this relationship is not so obvious. For example, in terms of NSE, Figure 6 (top left) shows low performances for dry (B4, B10, B11, B12) and wet catchments (B5, B3). Besides, Figure 8 suggests that post-processors methods are more reliable for wet basins (B1, B2, B3), but we also find unreliable results for B5 and B7, which are also wet catchments. Therefore, we do not see a clear relationship between the performance of the post-processing methods and the aridity index.

To illustrate these results, a monthly streamflow prediction time series at Guadalupe catchment (B12) is displayed in Figure 9. This catchment was selected as it the driest catchment of the MOPEX database. As seen in Figure 9, post-processed streamflow predictions produce a realistic and reliable uncertainty band, which contains all the observations and maximise precision without sacrificing reliability (Gneiting et al., 2007). Again, GMM post-processing produces the narrowest predictive uncertainty band.

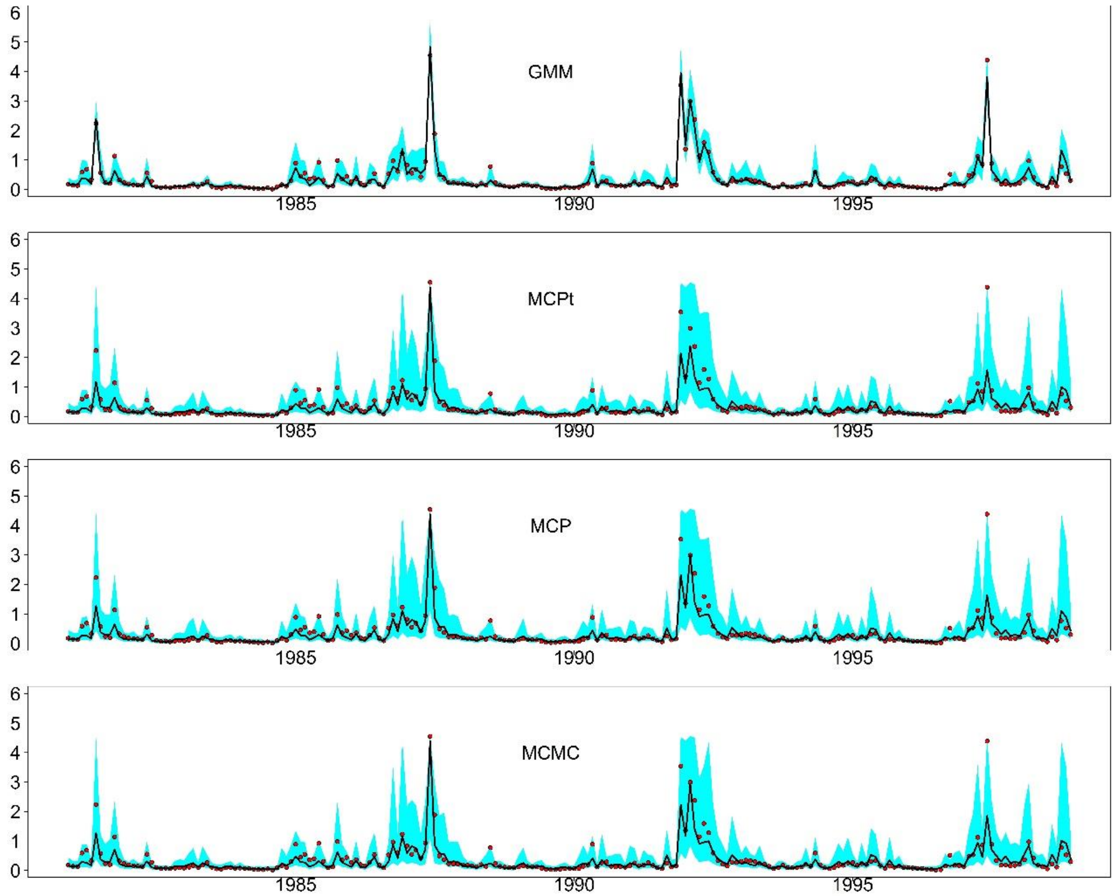

Figure 9. Conditional predictive uncertainty of monthly streamflow time series (m3 s-1) at the Guadalupe catchment (B12) during the validation period (1980-1998) by four post-processing methods. Red dots indicate observations, and black lines indicate predictive.

Overall, Figure 10 indicates that the GMM post-processing is superior to all other methods for the majority of the verification metrics. This relative superiority can be attributed to Gaussian mixture models (GMM) provide great flexibility for estimating the joint probability distribution, and each component of a Gaussian mixture density is associated with a group or cluster, i.e. low, medium and high streamflows. Nevertheless, the use of Gaussian mixture models is not new in hydrological post-processing. For example, Feng et al. (2019) recently used the

Gaussian mixture for estimating the marginal distribution of the hydrological uncertainty processor (HUP), and Klein et al. (2016) that applied Gaussian mixture for estimating the marginal distribution of the Pair-Copula uncertainty processor (COP). Note that Gaussian mixture was used for estimating the marginal distribution, but we used it for modelling the joint probability distribution that describes the relationship between deterministic hydrological predictions (predictors) and corresponding observed streamflows (predictands). Moreover, Figure 10 shows considerable overlap in the boxplots corresponding to MCP, MCPt and MCMC post-processing methods. This results suggest little difference in the performance of the post-processing methods.

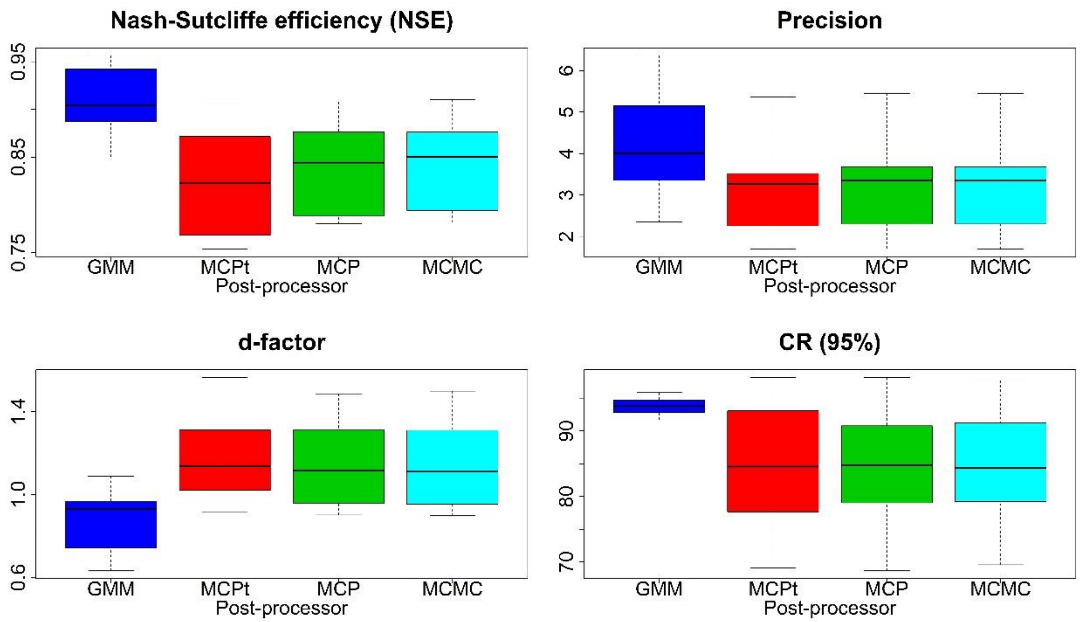

Figure 10. Performance indices of four post-processors during the validation period (1980-1998) averaged over all catchments.

2.4 Conclusions

In this research, we propose a new statistical post-processing method that combines the Bayesian joint probability (BJP) modelling approach and the

Gaussian mixture models (GMM) to quantify the conditional predictive uncertainty in monthly streamflows. We also compare the performance of the propose post-processor, which is called GMM post-processor, to existing Bayesian post-processors including the Model Conditional Processor (MCP), the MCP using the truncated Normal (MCPt) and the linear regression post-processor using Markov Chain Monte Carlo (MCMC) to inference parameters. To ensure the broad pertinence and generality of our conclusions, the assessment is applied to 12 MOPEX catchments. Three important conclusions are reached as follows:

1. The evaluation of uncorrected streamflow predictions via NSE shows the accuracy of the GR4J model deterministic predictions (NSE > 0.78). In general, we see small improvements as a result of the post-processing judging only by NSE. However, post-processors provide predictive distribution, which is essential for supporting robust decisions on environmental and water resources management.
2. From the post-processing methods considered in this study, the GMM post-processor is found to be the best suited for monthly streamflow predictions. The GMM post-processor delivers the sharpest predictive uncertainty without sacrificing reliability. However, the GMM approach has the disadvantage of any data-driven model, requiring a large number of fitting parameters.
3. MCP, MCPt and MCMC post-processing methods substantially perform similarly for monthly streamflow predictions in terms of all verification metrics evaluated (Figure 10).

Further work is needed to extend the GMM post-processor to handle multiple inputs, which is a multivariate model approach, multiple lead-time and multi-temporal approach, such as Coccia and Todini (2011). Future comparison studies are also necessary to evaluate the GMM post-processor with different hydroclimatic variables (i.e., soil moisture, water level, precipitation, temperature, etc.) and different temporal scale (i.e., hourly, daily, weekly, etc.). More work is necessary to advance in post-processing extreme data, zero flows and missing data. Besides we need post-processing methods for handling non-stationary conditions.

Rigorous predictive uncertainty quantification is increasingly viewed as essential in hydrological predictions (Kavetski and Clark, 2010). Therefore, a significant practical application of this research is the development of a robust monthly streamflow predictions post-processing approach that quantifying the predictive uncertainty of deterministic model outputs. Finally, the improvements in post-processed streamflow predictions achieve using the GMM post-processor will support water managers to make robust decisions in environmental and water resources management applications.

CHAPTER 3. Hydrological post-processing based on approximate Bayesian computation (ABC)

3.1 Introduction

Making unbiased, accurate and reliable streamflow predictions has regularly been one of the main goals for hydrologists. These hydrological predictions are valuable for risk assessment, water resources management, and ecological issues. Characterizing, quantifying, reducing and communicating the uncertainty of predictions is essential for decision-making and water management under anthropogenic conditions (Butts et al., 2004; Liu and Gupta, 2007). Uncertainty is everywhere and is impossible to avoid it (Lindley and Smith, 1972). Indeed, uncertainty is a fact of hydrology (Wilby and Harris, 2006). Generally speaking, uncertainty analysis is a crucial part of hydrological modelling process (Schoups and Vrugt, 2010). It is particularly useful for modelling comparison and selection (Schoups et al., 2008), improving model predictions, supporting decision-making (Reichert et al., 2015) and advancing towards reliable measurement systems. Most significantly, uncertainty analysis plays a considerable role in applications and the dialogue with decision-makers (Montanari and Koutsoyiannis, 2012). Some sources of uncertainty include input errors (e.g. rainfall sampling, low gauge density, interpolation method), epistemic errors (e.g. model parameters, model structure), and output errors (e.g. associated with rating curve errors). The propagation of confidence bounds from different uncertainty sources to model output is crucial for hydrologic modelling (Liang et al., 2012). Applying statistical post-processing methods is a useful approach to quantify the joint effect of these uncertainties. Hydrologic post-processors are statistical models that relate observed variables of interest (streamflow, water level) to predictors derived from deterministic hydrologic model outputs (Ye et al., 2014). The hydrologic post-processing aim is to reduce biases and quantify the uncertainty of deterministic predictions (W. Li et al., 2017).

In the context of conditional predictive uncertainty, which means fixed hydrological predictions, several techniques have been developed to quantify total uncertainty. Early works included methods as Model Output Statistics (MOS) (Glahn et al., 1972) and Hydrological Uncertainty Processor (HUP) (Krzysztofowicz and Kelly, 2000). More recent literature used the Bayes's theorem-based methods, such as Bayesian Model Average (BMA) (Raftery et al., 2005; Vrugt and Robinson, 2007), Model Conditional Processor (MCP) (Coccia and Todini, 2011; Todini, 2008) and Bayesian Joint Probability (BJP) (Wang et al., 2009). Moreover, there exists a variety of regression-based models, including a meta-Gaussian approach (Montanari and Brath, 2004; Montanari and Grossi, 2008), quantile regression (Weerts et al., 2011) and General Linear Model Post-Processor (GLMPP) (Zhao et al., 2011). Also, many other methods have been proposed, including non-parametric post-processor (Brown et al., 2010), machine learning (Solomatine and Shrestha, 2009), data-driven resampling techniques (Sikorska et al., 2015), Bayesian neural networks (Zhang and Zhao, 2012) and post-processing with error model (Evin et al., 2014). Several copula models have been proposed like a BMA-copulas (Madadgar and Moradkhani, 2014), pair-copulas in a multi-model ensemble (Klein et al., 2016) and ensemble copula coupling (Schefzik et al., 2013).

Although there are many approaches to improve hydrologic predictions by reducing uncertainties, they have not been standardized (B. Li et al., 2017; Montanari and Koutsoyiannis, 2012; van Oijen, 2017; Wagener and Gupta, 2005), and less attention has been paid in presence of intractable likelihood. The Approximate Bayesian Computation (ABC) (Fenicia et al., 2018; Kavetski et al., 2018; Nott et al., 2012; Vrugt and Sadegh, 2013) method deals with inferential problems with intractable likelihood. By intractable likelihood, we mean that the likelihood function is unavailable in closed form or by numerical derivation (Robert, 2016). In this context and for the sake of simplicity, we use a monthly error model that is useful for water resources management applications. To our best knowledge, up to now, this is the first study that has proposed a hydrological post-processor based on approximate Bayesian computation (ABC). This study introduces a method to quantify the conditional predictive uncertainty in hydrological post-processing contexts when it is cumbersome to calculate the likelihood (intractable likelihood).

Sometimes, it can be difficult to calculate the likelihood itself in hydrological modelling, specially working with complex models or with ungauged catchments. Therefore, we propose the ABC post-processor that exchanges the requirement of calculating the likelihood function by the use of some sufficient summary statistics and synthetic datasets. The aim is to show that the conditional predictive distribution is qualitatively similar produced by the exact predictive (MCMC post-processor) or the approximate predictive (ABC post-processor). We test the ABC post-processor in two scenarios: i) The Aipe catchment (poor predictions) and ii) the Oria catchment (good predictions). Deterministic and probabilistic verification frameworks are used to compare the performance of the ABC post-processor with the Markov Chain Monte Carlo (MCMC) approach (Gelman et al., 2013), that works when the likelihood is tractable. The rest of the chapter is structured as follows. The theory and methods are described in the

Section 3.2, applications in the Section 3.3, followed by discussion and conclusions in Section 3.4.

3.2 Theory and methods

Biased, inaccurate, and unreliable predictions in hydrology are mainly consequence of several sources of uncertainties. A hydrologic post-processor is an approach to deal with uncertainties from deterministic hydrologic model outputs propagated from all upstream sources. Applying statistical post-processing methods is useful to quantify these uncertainties. In this section, we describe the theory of hydrologic post-processing focussing both on algorithms dealing with intractable likelihood (ABC post-processor) and tractable likelihood (MCMC post-processor). These two post-processors are compared through verification metrics to assess their performance.

3.2.1 Monthly streamflow post-processor

Let $\boldsymbol{y^s} = (y_1^s, \cdots, y_T^s)^T$ be the output of a deterministic hydrological model and $\boldsymbol{y^o} = (y_1^o, \cdots, y_T^o)^T$ the observations. The hydrologic post-processor works by relating model outputs (e.g., streamflow) to corresponding observations through a statistical model (Ye et al., 2014). It serves the purpose of removing model biases from all upstream uncertainty sources. In this chapter, we assume a linear model between $\boldsymbol{y^o}$ and $\boldsymbol{y^s}$

$$\boldsymbol{y}^o = \beta_0 + \beta_1 \boldsymbol{y}^s + \boldsymbol{\varepsilon}\,, \tag{16}$$

where vectors $\boldsymbol{y}^o$ and $\boldsymbol{y}^s$ are expressed in $m^3 s^{-1}$, β_0 and β_1 are statistical parameters, and $\boldsymbol{\varepsilon}$ is a random variable that represents the error term in the statistical model. The three components on the right-hand side symbolise three distinct sources of estimation errors. The first term, β_0, represents constant deviation and can be called the displacement error. The second term, with an error parameter β_1, denotes a scale error or dynamic-range error. The third term is a random error, which is assumed as independent and identically distributed, with zero mean value and a standard deviation of σ. The first two terms describe a deterministic relationship between $\boldsymbol{y}^o$ and $\boldsymbol{y}^s$, and they characterise the systematic or epistemic error with β_0 and β_1. The error term expresses random fluctuations due to the effect of factors out of our control or measurement, and it is assumed that $\boldsymbol{\varepsilon_i} \sim N(0, \sigma^2)$ identical independent distributed (i.d.d.). This assumption is the most common, but it can be relaxed in favour of more general cases (e.g. Schoups and Vrugt (2010)). The linear model and the assumption of normality on the error term can be too restrictive for hydrological post-processors. Nevertheless, the linear error model is our first approximation although it is not always appropriate. We know that some circumstances require non-linear error models, but this linear error model worked for our water resources management application. In fact, Tian et al. (2016) proved that a linear model and three parameters are sufficient to fully capture the characteristic error of monthly predictions. The ABC is not strongly influenced by the model and the assumptions, since it is based on a distance measure between summary statistics of the observed and simulated data. This point will be clarified in the Section 3.2.3. To complete the error assumptions $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$, the observed and simulated streamflows are transformed to the Normal space previously applying the Normal Quantile Transformation procedure (NQT). Waerden (1953) described the theory behind the NQT, and Krzysztofowicz and Kelly (2000) demonstrated its application in hydrology. Figure 11 shows the steps we used to derive the conditional predictive uncertainty distribution.

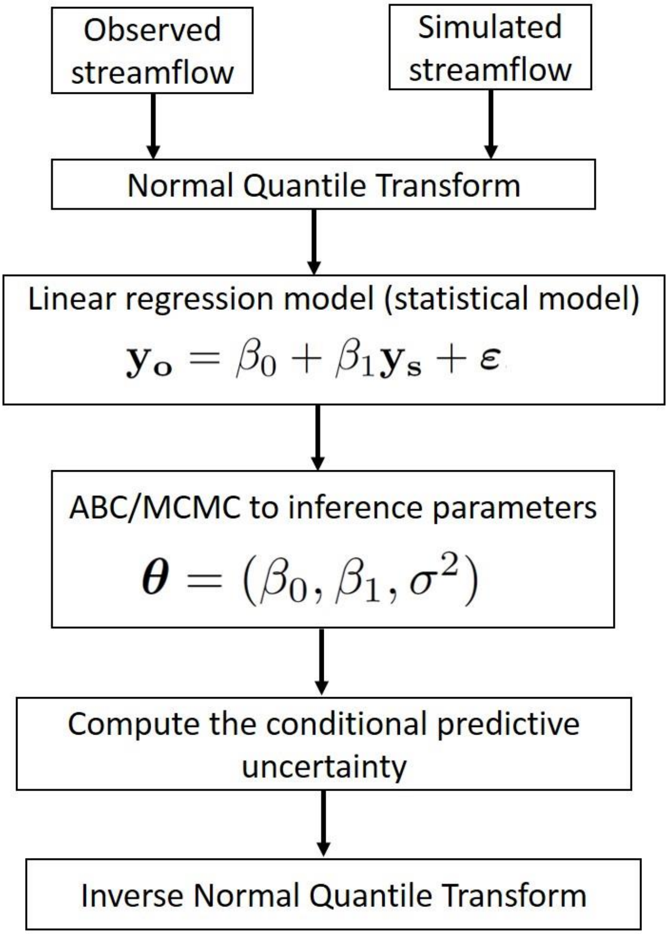

Figure 11. Flow chart of the process we used to derive the conditional predictive uncertainty distribution.

3.2.2 MCMC for monthly streamflow post-processor in Bayesian framework

In Bayesian Statistics, parameters are treated as random variables and inference is based on the posterior parameter distribution. The posterior parameter distribution of model (16), can be written, from the Bayes theorem, as:

$$p(\boldsymbol{\theta}|\boldsymbol{y}^o, \boldsymbol{y}^s) = \frac{p(\boldsymbol{y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s)p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s)p(\boldsymbol{\theta})\ d\boldsymbol{\theta}}, \tag{17}$$

where $p(\boldsymbol{\theta})$ indicates the prior parameter distribution, $p(\boldsymbol{y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s)$ denotes the likelihood of $\boldsymbol{y}^o$ conditional on the parameters $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2)$ and the deterministic output $\boldsymbol{y}^s$. We assume flat uniform priors for $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2)$ and from the assumptions on model (16) it follows that $\boldsymbol{Y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s \sim N(\mu = \beta_0 + \beta_1 \boldsymbol{y}^s, \sigma^2)$. Given the model it is not possible, as most of the time, to compute in closed form the integral in the denominator of the equation (17). So, we approximate the posterior distribution (17) by using the MCMC algorithms (Gelman et al., 2013). Specifically, we use the adaptive Metropolis-Hastings (Haario et al., 2001) to perform the Bayesian inference. This algorithm has been shown to perform adequately in hydrologic problems (Marshall et al., 2004). For detail on the implementation of adaptive Metropolis-Hastings algorithm to hydrologic modelling studies, relate to the research stated by Marshall et al. (2004).

In this thesis we are most interested in conditional prediction uncertainty. Let $\widetilde{y^o}$ be a future observation for model (16), then the posterior predictive density (which incorporate our uncertainty) of a future observation $p(\widetilde{y^o}|\boldsymbol{y}^s)$, is given by

$$p(\widetilde{y^o}|\boldsymbol{y}^s) = \int_{\emptyset} p(\widetilde{y^o}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{y}^o, \boldsymbol{y}^s)\, d\boldsymbol{\theta} \tag{18}$$

In words, the posterior predictive density is an average of conditional predictions over the posterior distributions of parameters (Gelman et al., 2013), reflecting both the uncertainty of the model and the uncertainty due to variability in future observations (Yoon et al., 2010).

3.2.3 ABC post-processor

The idea behind the ABC approach was introduced in population and evolutionary genetics by Pritchard et al. (1999) and Tavaré et al. (1997). Furthermore, Nott et al. (2012) were the first to introduce the ABC method in hydrology community. ABC is adequate for inference problems where sampling from the assumed probability model is much easier than evaluating its probability density function (e.g., intractable likelihood) (Fenicia et al., 2018). Using the ABC does not evade the requirement of Bayesian inference to stipulate a probability model of the data, but rather exchanges the requirement of calculating the likelihood function by the requirement of sampling model output realizations (Kavetski et al., 2018). ABC is a class of sampling methods that bypass exact likelihood calculations with a simulation of the model that produces synthetic datasets. The method then relies on some metric (distance) to compare simulated data to the data that were observed (Turner and Van Zandt, 2012). Then, the aim is to obtain an estimate of the posterior distribution of the parameter of the model. Recall that the posterior parameter $\boldsymbol{\theta}$ is the distribution of that parameter conditioned on the observed data and the deterministic output. Without a likelihood, it is not possible to write down an expression for this posterior, or to estimate it using Monte Carlo methods. However, we can simulate data $\boldsymbol{y}_{sim}$ using some $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ and retain $\boldsymbol{\theta}^*$ as a sample from the posterior if some pre-defined distance $d\{\boldsymbol{y}_{sim}, \boldsymbol{y}^o\}$ between the observed and the simulated data is less than some small value ϵ_0. There are three main ABC algorithms: 1) accepted-rejected (Beaumont et al., 2002), 2) Markov Chain Monte Carlo ABC (Marjoram et al., 2003), and 3) Sequential Monte Carlo ABC (Sisson et al., 2007). Although some ABC algorithms are more efficient, whether the ABC algorithm is optimized (or not), is not relevant to the results; hence, we use the accepted-rejected algorithm of Beaumont et al. (2002) for our application. We developed the ABC post-processor in R (Core Team, 2013) using the package abc (Csilléry et al., 2012). The pseudo code for the ABC is summarized in Algorithm 1.

Algorithm 1 ABC accepted/rejected algorithm

```
1: Sample θ*_i, i = 1, ..., N from the prior: θ* ~ p(θ)
2: Generate data y^i_sim = (y^i_1, y^i_2, ..., y^i_T)^⊤, i = 1, ..., N, from the model, p(·|θ*_i)
3: For 0 ≤ i ≤ N, store θ*_i if:
       d{η(y^i_sim), η(y^o)} ≤ ε0
   where η(·) is a vector statistic, d{·,·} is a distance criterion, and, given N, the tolerance
   level ε0 is chosen to be small.
```

Algorithm 1 proceeds in the following way: first, we sample a candidate parameter value $\boldsymbol{\theta}^*$ from the flat prior distribution. We then use this candidate to simulate a dataset $\boldsymbol{y}_{sim}$ from the normal model of interest that has the same number of observations as the observed data $\boldsymbol{y}^o$. Thereafter, we compare the simulated data $\boldsymbol{y}_{sim}$ to the observed data $\boldsymbol{y}^o$ by computing a distance between them given by a distance function $d\{\eta(\boldsymbol{y}_{sim}), \eta(\boldsymbol{y}^o) \leq \epsilon_0\}$. For computational ease, it is often convenient to define $d = \{\cdot,\cdot\}$ as a distance between summary statistics $S(\boldsymbol{y}_{sim})$ and $S(\boldsymbol{y}^o)$. Ideally, the summary statistics $S(\cdot)$ should be sufficient for the parameter $\boldsymbol{\theta}$. In this study, we consider five summary statistics including the sample mean, variance, skewness, kurtosis, and first sample autocorrelation (Fearnhead and Prangle, 2012) and run the algorithm 1 using the Euclidean distance between summary statistics and tolerance level $\epsilon_0 = 0.01$. Algorithm 1 produces the empirical posterior parameter distribution which we indicate with $p^*_{\epsilon_0}\big(\boldsymbol{\theta}|\eta(\boldsymbol{y}_{sim}),\eta(\boldsymbol{y}^o)\big)$. Estimates of the parameter $\boldsymbol{\theta}$ can be obtained by calculating the mean, mode or median of this empirical distribution. However, our interest in the chapter is not on these estimates but on the predictive posterior uncertainty. The predictive posterior uncertainty is formally defined as

$$g\big(\widetilde{y^o}|\boldsymbol{y}^s\big) = \int_{\emptyset} p\big(\widetilde{y^o}|\boldsymbol{\theta},\boldsymbol{y}^o,\boldsymbol{y}^s\big)p(\boldsymbol{\theta}|\boldsymbol{y}^o)\, d\boldsymbol{\theta}\,, \tag{19}$$

and it is approximated by

$$g^*\big(\widetilde{y^o}|\boldsymbol{y}^s\big) = \frac{\mathbf{1}}{\boldsymbol{M}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{M}} p\big(\widetilde{y^o}|\boldsymbol{\theta}^*_i,\boldsymbol{y}^o\big), \tag{20}$$

where $\boldsymbol{M}$ is the number of retained $\boldsymbol{\theta}^*$. It has been shown that even though the posterior parameter predictive distribution $p^*_{\epsilon_0}\big(\boldsymbol{\theta}|\eta(\boldsymbol{y}_{sim}),\eta(\boldsymbol{y}^o)\big)$ is not close to the true posterior distribution in (17) the posterior predictive distribution $g^*\big(\widetilde{y^o}|\boldsymbol{y}^s\big)$ under some regularity conditions can still be valid to approximate the (19) (Marin et al., 2012). In particular, there are three main regularity conditions: first, the data generating process (DGP) is correctly specified (probability model); second, both $p^*_{\epsilon_0}\big(\boldsymbol{\theta}|\eta(\boldsymbol{y}_{sim}),\eta(\boldsymbol{y}^o)\big)$ and $p(\boldsymbol{\theta}|\boldsymbol{y}^o,\boldsymbol{y}^s)$ are Bayesian consistent for the true value of $\boldsymbol{\theta}$ and large samples. Blackwell and Dubins (1962) and Diaconis and Freedman (1986) proved that, if all three regularity conditions are satisfied, then the $g\big(\widetilde{y^o}|\boldsymbol{y}^s\big)$ and $p\big(\widetilde{y^o}|\boldsymbol{y}^s\big)$ yield the same forecasting asymptotically. In other words, $g\big(\widetilde{y^o}|\boldsymbol{y}^s\big)$ and $p\big(\widetilde{y^o}|\boldsymbol{y}^s\big)$ merge asymptotically. Frazier et al. (2019) recently demonstrated theoretically and numerically the previous conclusion in economic models. Finally, the motivation for the use of ABC in hydrological models is evident: in cases where the likelihood is not accessible, the parametric posterior distribution itself is inaccessible and the integral in (18) cannot be computed via the MCMC methods.

3.2.4 Verification measures

Deterministic and probabilistic verification frameworks are used to assess outputs from the proposed ABC post-processor and the MCMC approach. We examine the accuracies, reliability, and robustness of the proposed method. These verification metrics are analysed during both the calibration and validation periods. In general, uncertainty analysis methods could be portrayed by the 95% uncertainty band that has to be as narrow as possible but still containing the largest amount of observations. Since verification metrics have been presented in the section 1.3, only a brief description of each is presented here.

As deterministic metrics, we include the Nash-Sutcliffe Efficiency (NSE), and the Kling-Gupta Efficiency (KGE) indices. NSE has been extensively applied to assess hydrological models. Likewise, KGE was presented as the modified version of NSE by Gupta et al. (2009). This metric involves the correlation, bias, and variability. Both NSE and KGE can range from $-\infty$ to 1 with NSE or KGE = 1 as a perfect fit between observation and simulation. As probabilistic metrics, we include the reliability and precision. Reliability refers to the statistical consistence of predictions with observed data, and precision refers to the concentration of the predictive distribution (small uncertainty). Zero is the worst reliability value while one is the best. To evaluate the reliability and precision of predictive distributions, Laio and Tamea (2007) suggested the use of PQQ plots (Thyer et al., 2009). In the PQQ plot context, if the predictive distribution and observed data are consistent, the corresponding p-value distribution should be uniformly distributed over the interval [0,1]. We apply the Kolmogorov-Smirnov test (K-S) to check this uniformity.

Finally, to gain more insight into the probabilistic metrics and following Li et al. (2017), we also compute the containing ratio (CR), which is the percentage of the measurement bracketed by this band, the average bandwidth of 95% uncertainty band (B) and the average deviation amplitude (D). We use the 95% prediction interval based on the 2.5 and 97.5 percentiles. As a result, an adequate predictive uncertainty is achieved when the CR is close to 95%. The smallest values of B and D are preferred. These three indices quantify the degree of predictions deviating from observations. Our strategy to compute comparative performance metrics is compatible with others similar studies such as Shafi et al. (2014), Ye et al. (2014), Khajehei and Moradkhani (2017).

3.3 Applications

Both methods are applied to monthly streamflows in two scenarios: the Aipe catchment with poor predictions (Colombia) and the Oria catchment with good predictions (Spain). By poor and good predictions, we mean that the $NSE < 0.5$ and $NSE > 0.8$ respectively. These scenarios allow a contrast in hydrology. The essential hydrologic features of both scenario catchments are summarised in Table 2.

Table 2. Hydrologic features of the two case studies catchments.

Catchment	Area (km^2)	P (mm/year)[a]	PET (mm/year)	Q (mm/year)	Run-off ratio (Q/P)	Aridity (PET/P)
Aipe	688.9	1922.71	1981.54	706.89	0.377	1.031
Oria	73	1498	733.4	765	0.511	0.489

[a] P mean areal precipitation, PET potential evapotranspiration and Q streamflow

The first scenario (poor predictions) is the Aipe river catchment in Huila State, in southern Colombia. Precipitation, potential evaporation and streamflow monthly time series are available from 1992 to 2012. The first fourteen years of data are used for model calibration, while the last six years served as a validation dataset to assess predictive capability. We used the abcd water balance model to simulate streamflows (Thomas, 1981). This model is a well-known conceptual spatially-lumped rainfall-runoff model which transforms precipitation and potential evapotranspiration data to streamflow at the catchment outlet. The hydrological model is selected for its conceptual simplicity and general usage. Figure 12 represents the output from the hydrological model. The abcd water balance model implemented in Aipe catchment is described in detail by Romero-Cuéllar et al. (2018).

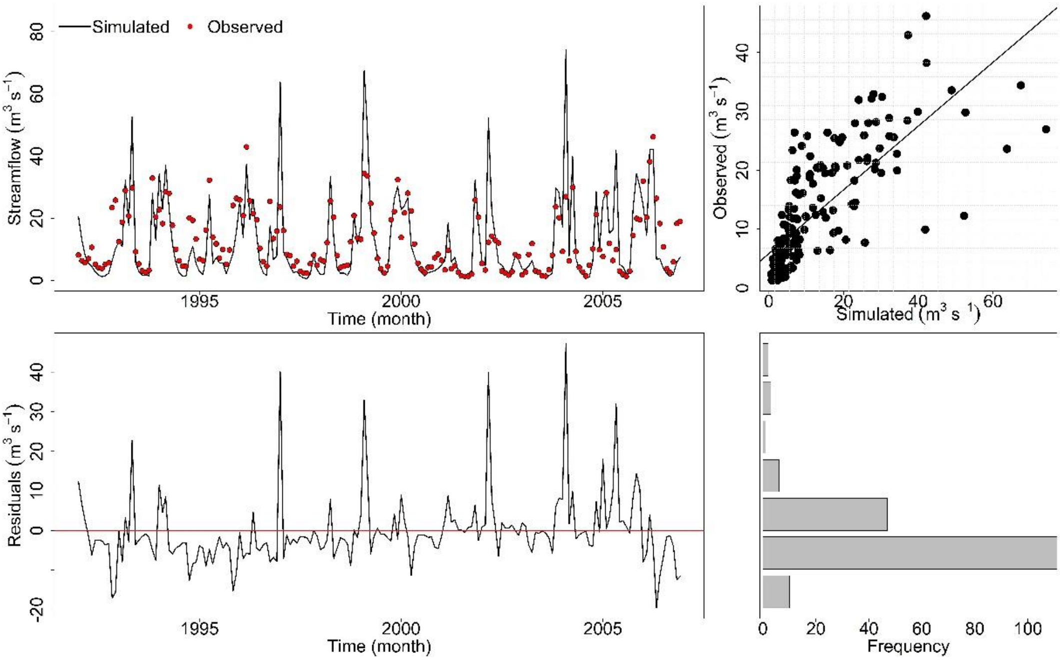

Figure 12. Monthly time series of deterministic streamflow predictions (solid line) and observations (red dots) from the Aipe river catchment. Scatter plot of simulated versus observed streamflows. Time series and histogram of the residuals.

The simulated streamflow time series in Figure 12 shows that deterministic hydrological predictions overestimate the observed streamflows. Moreover, the residuals time series shows high values, and the histogram of errors indicates a negative bias. The second scenario (good predictions) is the Oria river catchment located in the Basque Country Region, in northern Spain. The hydrological data for this catchment were collected from 1987 to 2000. We used ten years (i.e., 1990-2000) for calibration process and three years for validation (i.e., 1987-1990). Daily runoff simulations from the TETIS model aggregated to monthly values are used to estimate the predictive uncertainty at the C2Z1 Agauntza gauge. The TETIS model is a conceptual spatially-distributed hydrological model where each grid cell represents a tank model with six tanks connected among them. TETIS is a grid-based model, which takes advantage of all the spatially distributed information available. More details about the TETIS model and Oria application are in Francés et al. (2007) and Vélez et al. (2009). Figure 13 represents the performance of the hydrological model in the Oria river.

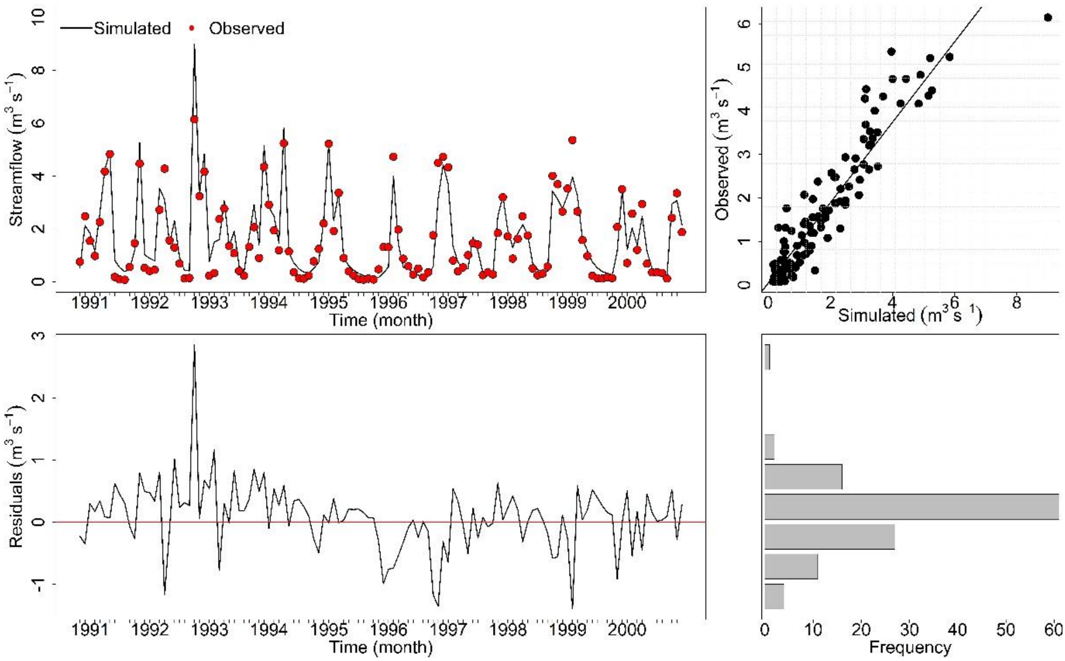

Figure 13. Monthly time series of deterministic streamflow predictions (solid line) and observations (red dots) from the Oria river catchment. Scatter plot of simulated versus observed streamflows. Time series and histogram of the residuals.

In contrast to the Aipe River, in the Oria River the hydrological model has better performance. The simulated streamflow time series in Figure 13 indicates that deterministic hydrological predictions correspond to the observed streamflows. Moreover, the error histogram is centred in zero. For each of the case studies, we calibrate a set of hydrological parameters. Calibrated parameters were achieved through optimisation which minimises the aggregated differences between simulated and observed streamflow values. Then, we use outputs from the hydrological model as inputs for the statistical model (hydrological post-processing). Next, the adaptative Metropolis-Hastings algorithm is implemented and run until the convergence of the parameter posterior distribution (statistical model) is achieved. Convergence is determined by both the visual trace plot evaluation of the posterior chains and the (Gelman and Rubin, 1992) R statistic which considers convergence in terms of the variance with a single chain and the variance between multiple parallel chains. Finally, we compute the conditional predictive distribution.

The results for monthly streamflow forecast are now presented. They are presented independently for both scenarios catchment and calibration and validation period. In section 3.3.1 the performance of the Aipe catchment is evaluated, while in section 3.3.2 the performance of the Oria catchment is assessed. For each of the scenarios, the MCMC algorithm is first applied to estimate the predictive uncertainty in time series domain, and the ABC algorithm is then used to assess the predictive uncertainty in summary statistic domain and free-likelihood function. As mentioned before, this chapter shows that the predictive superiority of the exact predictive (MCMC post-processor), over the approximate (ABC post-processor) using some sufficient summary statistics and synthetic datasets, is minimal. To achieve this aim, we use MCMC post-processor as a benchmark to make results more comparable with the proposed method.

3.3.1 First Scenario: The Aipe Catchment

Figure 12 represents a general view of hydrological model performance (deterministic predictions). Note that the hydrological model does not mimic observed data quite well, it does not reproduce maximum streamflow events. In other words, the model over-predicts peak streamflows. Moreover, Figure 12 shows that the error variance is heteroscedastic, and the error histogram is not normal. In summary, the hydrological model (deterministic predictions) has a poor performance. Regarding deterministic metrics, the most considerable improvement is the result of post-processing approaches (Table 3). For example, when post-processing is used, the NSE increases in as many as 74.63% for the calibration and 25.84% for the validation period. The acceptable range for NSE is considered to be above 0.5 (Moriasi et al., 2007); therefore, results in Table 3 point that the deterministic prediction does not have a satisfying skill as compared with the post-processing approaches, which frequently guides to valid performance metrics because they work directly to improve the errors in model outputs (Ye et al., 2014).Visible improvements are also inspected concerning KGE; for instance, KGE increases in as many as 30.3% for the calibration and 16% for the validation period (improvements are not as pronounced as for NSE).

Table 3. Deterministic and probabilistic performance metrics of the raw prediction, MCMC and ABC post-processor for the Aipe catchment.

Performance metric	Calibration			Validation		
	Deterministic predictions	Post-processing		Deterministic predictions	Post-processing	
		MCMC	ABC		MCMC	ABC
NSE	0.165	0.669	0.671	0.571	0.777	0.773
KGE	0.527	0.769	0.764	0.637	0.757	0.744
Reliability		0.996	0.996		0.993	0.993
Precision		2.403	2.306		2.581	2.5
K-S test (p-value)		0.465	0.75		0.132	0.223
B (m^3s^{-1})		14.95	15.64		25.78	26.86
CR (%)		88.33	88.89		94.44	95.83
D (m^3s^{-1})		6.82	6.92		12.23	12.42

Regarding probabilistic metrics, the predictive PQQ plot of the MCMC (upper-right) and ABC (lower-right) post-processing is presented in Figure 14 during the calibration period. Figure 14 shows realistic narrow predictive bounds because only some peak flows are not bracketed, and all low streamflows are bracketed. It is important to notice that the upper and lower predictive uncertainty appearing in the Figure 14 are almost identical. Besides, it is clear from these figures that the curves closely follow the bisector. This means that the predictive distributions of post-processed streamflows are reliable. We also see that reliability indexes are close to one, further confirming that predictions are reliable (Table 3).

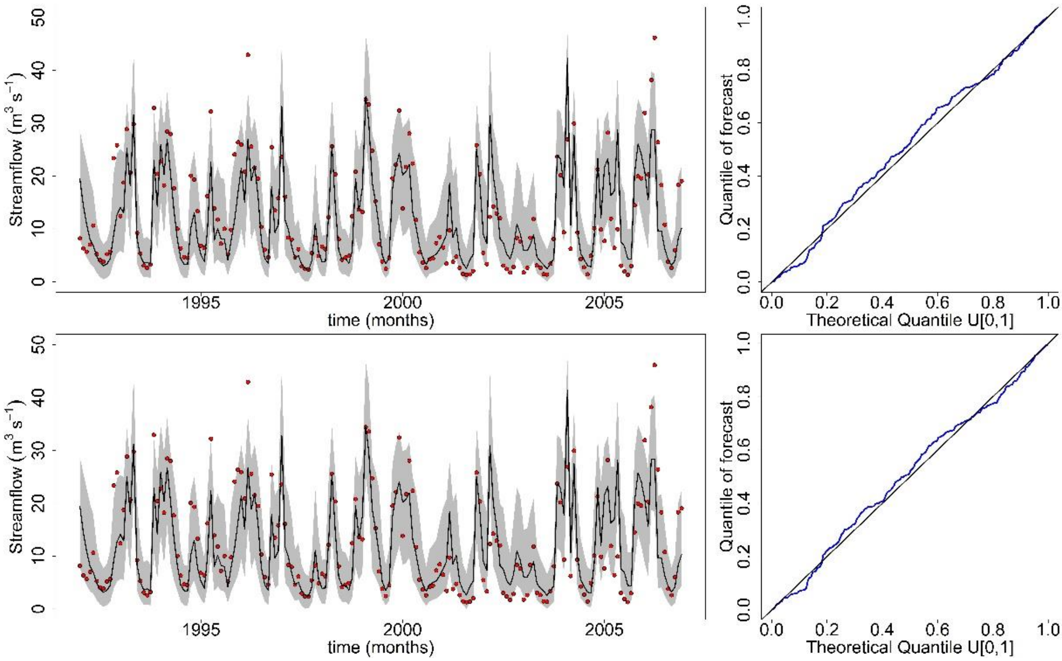

Figure 14. Conditional predictive uncertainty from MCMC (upper) and ABC (lower) post-processor on the Aipe catchment. PQQ-plot of the conditional predictive distribution (right). Dots indicate observations, line indicates median prediction and grey region indicates 90% uncertainty.

Generally speaking, we do not find crucial differences between the MCMC and ABC post-processing approaches. Reliability and precision metrics are practically equal for both post-processing methods in calibration and validation periods (Table 3). Additionally, both techniques pass the K-S test. In the calibration period, the B and D metrics of ABC post-processor are slightly higher than the MCMC post-processor. Table 3 further reports the coverage rate (CR) for testing the sharpness of the conditional predictive uncertainty. Perfect predictive distribution would expect that the CR close to the assumed 90% prediction level. In the calibration period, the CR for the two post-processors are quite similar. In general, all verification metrics of both methods deteriorate in the validation period except for the CR that improves slightly. These results suggest that the ABC post-processor which uses just some sufficient summary statistics and a free-likelihood function may have similar performance to the MCMC post-processor that uses a likelihood function. Therefore, the ABC post-processor has a satisfactory performance. The second scenario (section 3.3.2) investigates whether these findings would hold for good hydrological predictions.

3.3.2 Second Scenario: The Oria Catchment

In the second scenario, we analyse an oceanic climate and spatially-distributed hydrological model. In contrast to the first catchment, the hydrological model (deterministic predictions) has a good performance as seen in verification metrics in Table 4.

Table 4. Deterministic and probabilistic performance metrics of the raw prediction, MCMC and ABC post-processor for the Oria catchment.

Performance metric	Calibration			Validation		
	Deterministic predictions	Post-processing		Deterministic predictions	Post-processing	
		MCMC	ABC		MCMC	ABC
NSE	0.875	0.91	0.911	0.939	0.955	0.956
KGE	0.918	0.903	0.91	0.891	0.909	0.917
Reliability		0.995	0.995		0.982	0.982
Precision		2.95	2.87		2.28	2.19
K-S test (p-value)		0.972	0.923		0.868	0.872
B (m^3s^{-1})		1.47	1.51		1.34	1.37
CR (%)		86.07	86.07		77.78	80.56
D (m^3s^{-1})		0.85	0.86		0.66	0.67

In general, post-processing approaches improve performance forecasts. Nevertheless, improvements are not as pronounced as for the first scenario. For instance, when post-processing is used, the NSE increases barely 3.9% for the calibration and 1.8% for the validation period. Furthermore, we do not find improvements regarding the KGE. Concerning probabilistic metrics, Figure 15 shows time series and predictive PQQ plots for streamflows predictions using the MCMC (upper) and ABC (lower) post-processing during the calibration period. As is evident in the Figure 15, the exact (MCMC) and the approximate (ABC) conditional predictive uncertainty are seen to be an extremely close match. In particular, prediction uncertainty bands supply an adequate description of observed values. The predictive PQQ plots in Figure 15 confirm that predictions under post-processing are providing adequate representation of observed streamflows, as PQQ plots follow the bisector. This means that the predictive distributions of post-processed streamflows are reliable. Besides, we can confirm this by the reliability index (Table 4).

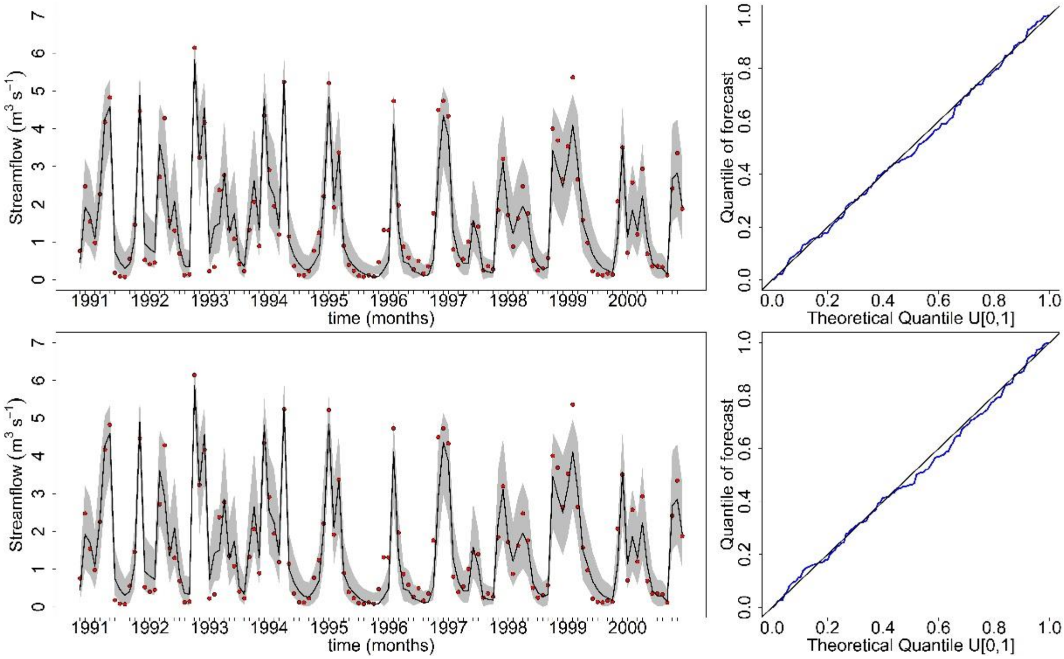

Figure 15. Conditional predictive uncertainty from MCMC (upper) and ABC (lower) post-processor on the Oria catchment. PQQ-plot of the conditional predictive distribution (right). Dots indicate observations, line indicates median prediction and grey region indicates 90% uncertainty.

As is consistent with the first scenario, the approximate (ABC) conditional predictive uncertainty is almost equivalent to the exact predictive (MCMC) (Figure 15). Reliability and K-S test metrics are practically equal for both post-processing approaches in calibration and validation periods. Only a tiny difference is identified in the precision metric (Table 4). This finding is not surprising as the ABC uses summary statistics. Regularly summary statistics are linked with loss of information. This loss of information may be undesired for predicting objectives.

Nevertheless, this problem could be avoided by using a set of sufficient summary statistics. In particular, during the calibration and validation periods, the B and D metrics of ABC post-processor are slightly higher than the MCMC post-processor, indicating wider predictive uncertainty bounds than the MCMC post-processor. For this reason, ABC method has 2% more of the observed samples in the 90% predictive uncertainty than the MCMC post-processor during the validation period (Table 4). Finally, a performance comparison between the scenarios based predictive uncertainty shows that, the scenario 2 gave narrower prediction uncertainty bands than scenario 1. This is due to the quality of hydrological predictions influences the conditional predictive uncertainty. Contrary to scenario 1, the CR for both post-processors deteriorate in the validation period. To sum up, in both scenarios there is little visual distinction between the approximate (ABC post-processor) and the exact (MCMC post-processor) conditional predictive uncertainty.

3.4 Discussion and Conclusions

The main aim of this chapter is to show that the conditional predictive distribution is qualitatively similar produced by the exact predictive (MCMC post-processor) or the approximate predictive (ABC post-processor) using some sufficient summary statistics and synthetic datasets. To achieve this aim, we use MCMC post-processor as a benchmark to make results more comparable with the proposed method. We apply both methods (ABC and MCMC) to two scenarios: the Aipe catchment (Colombia) and the Oria catchment (Spain). The advantage of the proposed method (ABC post-processor) is highlighted in the inferential problems with intractable likelihood. Sometimes, it can be difficult to calculate the likelihood itself in hydrological modelling, specially working with complex models or with ungauged catchments.

As well as Ye et al. (2014), Bogner et al. (2016) and Woldemeskel et al. (2018) we confirm that post-processing techniques can improve forecasts significantly when hydrological model predictions are especially poor (Table 3). Furthermore, we find through our numerical evidence that the MCMC and ABC post-processors provide similar predictive performance. This result was confirmed by Fenicia et al. (2018) and Kavetski et al. (2018), that the shape of the predictive distribution is qualitatively similar produced by the exact predictive (MCMC) or approximate predictive (ABC). These findings can be attributed to the correct specification of the probability model (data generating process). When the assumed probability model is correctly specified, just a little information is lost regarding predictive performance. In contrast, significant differences can appear when the probability model to represent observed streamflows is inadequate. In addition, Frazier et al. (2019) demonstrated theoretically that the ABC made forecasts which were asymptotically similar to those obtained from the exact Bayesian methods when the sample was large, the data generating process was correctly specified and, if the conditions for Bayesian consistency and asymptotic normality of both the exact (MCMC) and the approximate (ABC) posteriors were satisfied. Moreover, we recommend to check Drovandi and Pettitt (2011) for the use of ABC with synthetically case studies in the presence of intractable likelihood.

In both scenarios, we have proved that the predictive superiority of the exact predictive (MCMC post-processor) over the approximate (ABC post-processor) is minimal for hydrological models. There are two significant differences between this study and the previous works (e.g. Fenicia et al. (2018) and Kavetski et al. (2018)). First, they used the ABC method to calibrate jointly hydrological models and to compute the predictive uncertainty. Instead of Bayesian calibration of hydrological models, we used the ABC method in hydrological post-processing context, so we calculated the conditional predictive uncertainty, while Kavetski et al. (2018) calculated the predictive uncertainty. Second, Fenicia et al. (2018) and Kavetski et al. (2018) did not use summary statistics, which is a tremendous difference with this study. In other words, we obtained a similar main conclusion but in a different context and method.

Although we used a comparative analysis between MCMC and ABC hydrologic post-processors, it should be noted though that this study does not aim to show that ABC post-processor has a better performance than MCMC post-processor. Thus, we use MCMC post-processor as a benchmark to make results more comparable with the proposed post-processor. Moreover, we know that the MCMC method has proven its ability in hydrological predictions. In addition, it is true that any comparison can be affected by different factors, but it must be emphasised the focus of this comparison is between two post-processors without particular emphasis on the hydrological model performance. Actually, they are just predictions, which are the input for both post-processors in two scenarios (poor and good predictions).

We also applied the NQT transformation to achieve assumptions of the error model, but any transformation produces information loss. Besides, we know that our analysis is conditioned to the linear regression model (hydrologic post-processor). In standard linear regression, the average link between observed and simulated streamflows is summarised with a single slope parameter expressing this relationship. However, Diks and Vrugt (2010) pointed out that a simple regression method could result in improvements equivalent to more complex methods. Furthermore, our idea is to show that the ABC approach can be used to compute the conditional predictive uncertainty rather than to perform a complex post-processor, and therefore, we can tolerate some of the less realistic assumptions. Although the examples that we have used in this chapter are moderately simple, generally speaking, the ABC post-processor is highly flexible and can be used for more complex models. Future research should develop methods that relax any transformation, evaluate the impact of pre-processing and post-processing and explore the multi-regression model. We only scrutinize the ABC to compute the approximate conditional predictive uncertainty in hydrological post-processing context, but there are other approximations, for instance, Bayesian synthetic likelihood (Price et al., 2018), Bayesian empirical likelihood (Mengersen et al., 2013), variational Bayes (Tran et al., 2017) and bootstrap methods (Zhu et al., 2016).

The ABC post-processor has potential in areas such as operational hydrology, flood protection, drinking water production, risk assessment, irrigation management, water resources management and ecological issues. Besides, the ABC post-processor offers the opportunity to improve decision support with intractable likelihood function, e.g. to predict in ungauged basins or evaluate climate change predictive uncertainty. In summary, we conclude that deterministic predictions (no post-processing) perform poorly concerning deterministic verification metrics, and the MCMC and ABC post-processors provide similar predictive performance. Therefore, the approximate Bayesian computation may then be used as an alternative method to estimate the conditional predictive uncertainty of hydrological predictions with intractable likelihood.

CHAPTER 4. Estimating the uncertainty of streamflow statistics obtained from climate projections using the ABC post-processor

4.1 Introduction

Climate change, and its impacts, mitigation and adaptation, are central water challenges for scientists and engineers in the 21st century (Weaver et al., 2013). The most confident climate change projections show increased frequency and intensity of extreme events, and the intensity, duration and frequency of rainfall events are expected to change with global warming (Wild and Liepert, 2010). These projections could have negative consequences for the development of humanity, such as biodiversity loss, droughts, floods, and so forth. Recently, Broderick et al. (2019) projected more frequent, severe and tenacious droughts, more regular, widespread and dangerous floods, and more detrimental water pollution episodes. Given these projections, making robust decisions and designing strategies for the management and adaptation to climate change are essential needs in a global world's dynamics (Winsemius et al., 2014).

In any case, proper robust decisions about water resources planning and management require knowledge of the uncertainty associated with climate projections (for example, see Ray and Brown (2015), Vogel (2017) and Kundzewicz et al. (2018)). At the same time, climate change projections provide valuable information for water management and water resources planning. Nevertheless, climate change impact studies have been a notable challenge because they have multiple sources of uncertainty (Clark et al., 2016; Krysanova and Hattermann, 2017; Wilby and Harris, 2006). Uncertainties in the climate change context may arise from scenarios of future socio-economic development (Latif, 2011) and carbon emissions (Schenk and Lensink, 2007), general circulation models (GCM) (Murphy et al., 2004), regional climate models (RCM) or downscaling models (Stoll et al., 2011), bias correction methods (Maraun, 2016; Maraun et al., 2017) and natural climate variability (Deser et al., 2012; von Trentini et al., 2019). Understanding, characterising, reducing and communicating the uncertainty of climate change projections become crucial to support decision making and water planning under anthropogenic conditions (Butts et al., 2004; Liu and Gupta, 2007).

Broadly speaking, uncertainty is a universal concept with different meanings. In particular, the predictive uncertainty concept is often confused with model uncertainty (Mantovan and Todini, 2006; Todini, 2008). In this thesis, we focus on the predictive uncertainty, which characterises our best knowledge of future outcomes. In other words, the predictive uncertainty is the probability of predictand (discharge, water level, precipitation, etc.) conditional over all knowledge that we obtain by mechanistic models (Krzysztofowicz and Kelly, 2000). Therefore, the aim of predictions is to portrayal the uncertainty of predictands rather than the uncertainty of forecasting produced by hydrological models (Todini, 2008).

From a broader perspective, the *standard multi-model ensemble approach* is the most common method to deal with uncertainty. In the multi-model ensemble approach, the ensemble members of all the participating models are pooled into a single sample with equal weights (Min et al., 2009). This approach is widely used for the climate community on the continental and global scales (Christensen et al., 2010), and its application to climate change was recommended by the Intergovernmental Panel on Climate Change – IPCC (Adler and Hirsch Hadorn, 2014). Another version of the multi-model ensemble approach selects or rejects models by considering their performance in the historical baseline period. Although this method is a non-trivial assignment (Clark et al., 2016; Maraun et al., 2010), this approach is preferred by the hydrological community (Krysanova et al., 2018). In summary, the prevailing perception is that climate impact assessments based on the multi-model ensemble approach provide robust results to quantify the uncertainty of climate change projections (Krysanova and Hattermann, 2017).

However, recent studies state that the multi-model ensemble method cannot adequately describe uncertainty (Fatichi et al., 2016; Gao et al., 2019; Nearing and Gupta, 2018). Some possible reasons for this could be that this approach needs sufficiently large numbers of models and ensembles to span the range of possible physical representations (Boberg and Christensen, 2012), and a model consensus does not indicate reliability (Maraun et al., 2017). In addition, the multi-model approach assumes that all models are independent and equally plausible, which is a strong assumption that is hard to justify for climate models (Knutti et al., 2019). Consequently, new methods have emerged to overcome the weaknesses of the multi-model ensemble approach. For example, Steinschneider et al. (2012) presented a statistical framework to quantify the uncertainties of hydrologic response with climate change using Bayesian modelling. Clark et al. (2016) suggested the quantitative hydrologic storylines of climate change impacts. Nearing and Gupta (2018) recommended the information theory for quantifying the epistemic uncertainty of climate change. Pechlivanidis et al. (2018) applied the minimum redundancy concept to identify a representative subset in a large model ensemble. Gao et al. (2019) proposed the resampling change factor of meteorological variables based on their probability information to provide reliable climate change projections. Von Trentini et al. (2019) advocated quantifying the contribution of natural/model-internal variability on the total uncertainty of a multi-model ensemble. Finally, Padrón et al. (2019) proposed constraining the multi-model ensemble by the Markov Chain Monte Carlo (MCMC) method. Note that neither of previous studies has used a post-processing approach.

Another way of quantifying the total uncertainty of climate change projections is to follow a hydrological post-processing approach, which is a statistical model that relates the observed variables of interest to the simulated ones (W. Li et al., 2017). In a broader context, statistical post-processors are important tools for reducing systematic errors and obtaining an appropriate assessments of the predictive uncertainty (Buizza, 2018). Specifically, the statistical post-processor of the model outputs describes past prediction errors and uses this information to condition projections at a future time (Biondi and Todini, 2018). Currently, post-processing methods are well established in the weather forecasting community for short-term forecasting (Schepen et al., 2018), but forecasts and projections are conceptually different. Climate forecasts estimate climate evolution in the future, while climate projections depend on radiative forces scenarios. Therefore, climate projections are only a plausible state of future climate (Maraun et al., 2010). Moreover, climate change projections are not in synchrony with observations (Maraun, 2016). In other words, climate change projections are not expected to provide precise predictions of the temporal evolution of weather variables as seen in the observed time series.

The traditional climate change impact studies assume that a range of GCM-RCM chains properly represents the uncertainty of climate change projections. However, climate change projections are not in synchrony with observations (Maraun, 2016), so climate change projections model streamflow statistics instead of time series (Zhao et al., 2017). Unfortunately, a little information is available on the uncertainty of streamflow statistics. To bridge this gap, this study introduces a coherent univariate method, called ABC post-processor, to assess the conditional predictive uncertainty of streamflow statistics obtained from climate change projections that merge statistical post-processing, approximate Bayesian computation and streamflow statistics.

Romero-Cuellar et al. (2019) developed the ABC post-processor for inferential problems with intractable likelihood (see chapter 2), previously unavailable in a closed form or by numerical derivation (Robert, 2016). As the predictive uncertainty of climate change projections cannot be assessed by observations, but by summary statistics of observations, e.g. the mean monthly, these projections are a particular case of intractable likelihood problems. Although many studies on hydrological post-processing can be found, limited information is available on hydrological post-processing using approximate Bayesian computation and streamflow statistics to quantify the predictive uncertainty of climate change. In this chapter, we aimed to post-process climate change projected monthly streamflows and to assess their conditional predictive uncertainty by the ABC post-processor approach with 12 long-term mean monthly streamflows as the summary statistics for the baseline period (1987–2000). As an illustrative case study, we analysed the climate change projections (AR5 - IPCC) of the monthly streamflows in the upper Oria catchment (Spain), with deterministic and probabilistic verification frameworks to assess the ABC post-processor outputs. This framework can be used to provide essential water and environmental resources management information, to plan and operate water resource systems, and to support robust decisions about adaptation plans. The rest of the chapter is structured as follows: the methodology is described in Section 4.2, the case study in Section 4.3, the results in Section 4.4, and the discussion and conclusions in Section 4.5.

4.2 Methodology

4.2.1 General framework

To post-process climate change projected streamflows and to assess their conditional predictive uncertainty, we propose the general framework summarized in Figure 16. First of all, meteorological variables are downscaled for the baseline period. Secondly, a hydrological model is calibrated and validated using observed hydrometeorological data to follow the hydrological modelling procedure. Even if the objective is monthly streamflows, using a daily time discretisation is highly recommended to reduce the time scale effect of monthly simulations (Francés et al., 2007). Finally, the ABC post-processor is used to conduct the hydrological post-processing approach. In the following subsections, more detailed information is provided.

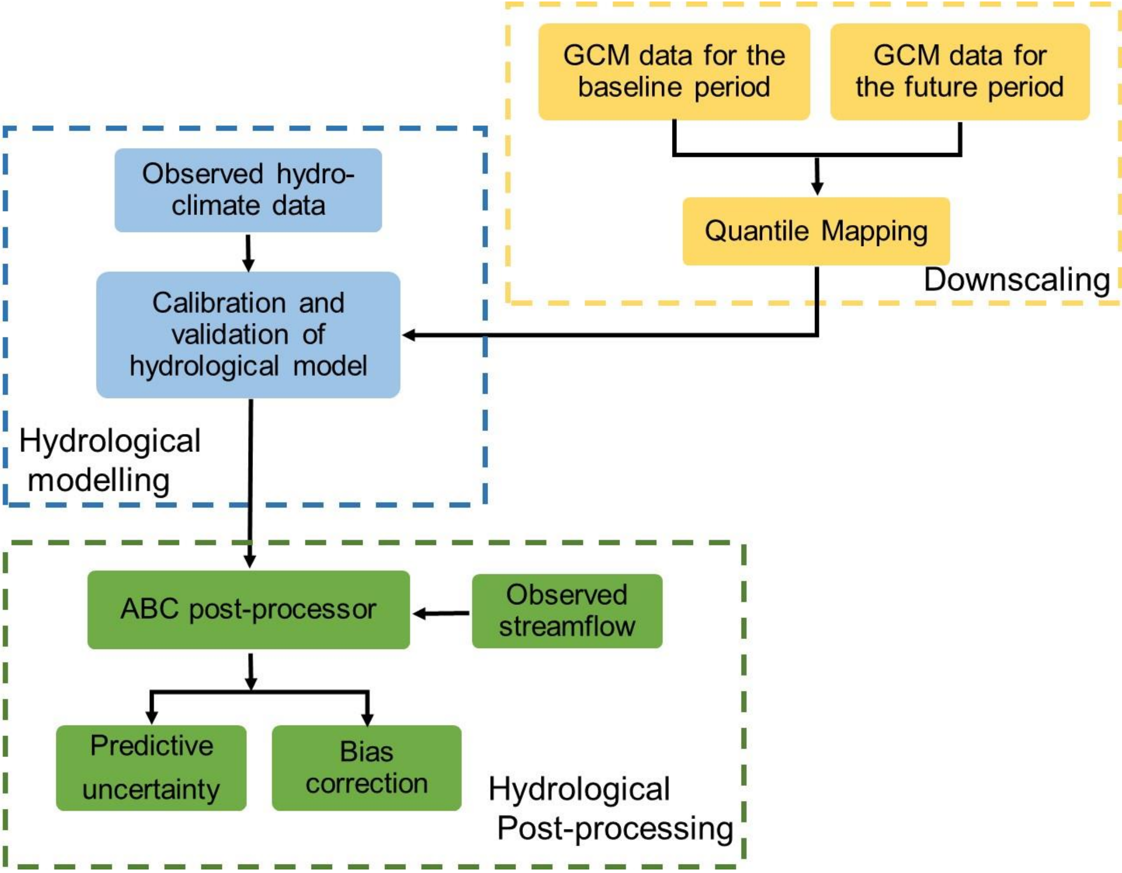

Figure 16. Framework of this study to post-process climate change projected streamflows and to assess their conditional predictive uncertainty.

The proposed framework is useful for evaluating climate models (GCM/RCM) and to constrain ensemble estimates of future streamflow changes to reduce the conditional predictive uncertainty, which is achieved by the ABC post-processor that uses the monthly mean streamflows as summary statistics and a free-likelihood approach.

4.2.2 The ABC post-processing procedure

The main characteristic of the ABC post-processor used herein is to outline what follows. For additional details on the ABC post-processor method, refer to Romero-Cuellar et al. (2019). The ABC post-processor is a Bayesian linear regression model that relates climate change projected streamflows (predictors) and observed streamflows (predictands):

$$\boldsymbol{y}^{o} = \beta_0 + \beta_1 \boldsymbol{y}^{s} + \boldsymbol{\varepsilon}, \tag{21}$$

where vectors $\boldsymbol{y}^{o}$ and $\boldsymbol{y}^{s}$ are the observed streamflows (predictands) and climate change projected streamflows (predictors), respectively; β_0 and β_1 are the model parameters and $\boldsymbol{\varepsilon}$ is an error term to account for random sampling noise. It is assumed that $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$ is identical independent distributed (i.i.d). As the ABC post-processor uses an aggregational approach to describe errors, all the sources of uncertainty are grouped into a linear residual error equation (21). The three components on the right-hand side denote three distinct sources of estimation errors. The first term, β_0, indicates constant deviation and can be called the displacement error. The second term, with an error parameter β_1, represents a scale error or dynamic-range error. The third term is a random error $\boldsymbol{\varepsilon}$, which symbolises aleatory uncertainty, and the first two terms characterise the systematic or epistemic error. It is known that Equation (21) and i.i.d assumptions are questionable (e.g., Sorooshian and Dracup (1980)), but the ABC post-processor applies the Normal Quantile Transformation (NQT) (Van Der Waerden, 1953) to normalised residuals. In addition, NQT moves to normal space using non-parametric distribution to estimate the marginal distribution of the random variables. Equation (21) is also adequate for water resources applications (Tian et al., 2016) and the ABC method is not strongly influenced by the model and his assumptions. The posterior distribution of the model (21) is, $(\boldsymbol{\theta}|\boldsymbol{y}^{o}, \boldsymbol{y}^{s})$, which is given by Bayes Theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{y}^o, \boldsymbol{y}^s\ ) = \frac{p(\boldsymbol{y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s)p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s)p(\boldsymbol{\theta})\ d\boldsymbol{\theta}}, \tag{22}$$

where $p(\boldsymbol{\theta})$ is a prior parameter distribution and $p(\boldsymbol{y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s)$ is a likelihood function. The ABC post-processor assumes flat uniform priors for $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2)$ and from the assumptions in model (1), it follows that $\boldsymbol{Y}^o|\boldsymbol{\theta}, \boldsymbol{y}^s \sim N(\mu = \beta_0 + \beta_1 \boldsymbol{y}^s,\ \sigma^2)$. The ABC post-processor approximates the posterior distribution (Equation (22)) using the accepted-rejected algorithm (Beaumont et al., 2002), which avoids the prerequisite of deriving the likelihood in a closed form and instead needs sampling from the underlying probability model and some sufficient summary statistics (Toni et al., 2009). The parameters of ABC post-processor parameters were estimated using the streamflow statistics of the climate change projected streamflows and observations as training samples during the baseline period (1987–2000). Specifically, the ABC post-processor computes the conditional predictive uncertainty, so let $\widetilde{y^o}$ be a future observation for the model (21), then the posterior predictive density of a future observation $p\left(\widetilde{y^o}|\boldsymbol{y}^s\ \right)$, is given by:

$$p\left(\widetilde{y^o}|\boldsymbol{y}^s\ \right) = \int_{\emptyset} p\left(\widetilde{y^o}|\boldsymbol{\theta}\right)p(\boldsymbol{\theta}|\boldsymbol{y}^o, \boldsymbol{y}^s)\ d\boldsymbol{\theta} \tag{23}$$

In other words, the posterior predictive density is an average of the conditional predictions over the posterior distribution (Gelman et al., 2013), by reproducing both the model's uncertainty and the uncertainty due to variability in future observed streamflows (Yoon et al., 2010). It is noteworthy that the ABC post-processor is a univariate model framework, which is one predictor (climate change model's output). Another assumption is that observations and climate change projections should be correlated.

In practice, for each climate models, the ABC post-processor was implemented with 100.000 simulations. The parameter values drawn from uniform prior distributions and the calculated associated streamflow statistics were used for each simulation. As streamflow statistics, we used the mean monthly streamflows for the baseline period (12 summary statistics). In this chapter, we implemented the accepted-rejected algorithm (Beaumont et al., 2002) using the Euclidian distance between summary statistics. A tolerance level was adaptively defined. It has been found that the tolerance level was 0.01 after a few tests. We developed the ABC post-processor in R (Core Team, 2013) using the package abc (Csilléry et al., 2012).

4.2.3 Verification metrics

Deterministic and probabilistic metrics were used to evaluate the outputs of the ABC post-processor and the 14 climate models during the baseline period. In particular, the accuracies, reliability, and robustness were examined. As verification metrics has been reported in the section 1.3, only a brief description of each is presented herein. The predictive performance of each climate model was assessed by comparing the 95% uncertainty band for a baseline period (1987–2000). The Nash-Sutcliffe (NSE) and Kling-Gupta (KGE) efficiency indices were also used to measure the predictive accuracy and the percentage of observations, which fell within the 95% prediction intervals to evaluate the reliability of the uncertainty band. The containing ratio (CR) assesses the percentage of observations bracketed by the 95% uncertainty band. In this study, a perfect uncertainty quantification was achieved when the CR came close to 95%. Precision rises along with the narrower uncertainty band and reliability, and increases the closer the 95% prediction level is reached. The visual test was run to check the predictive performance, as in previous studies (e.g., Najafi et al. (2011); Schoups and Vrugt, (2010); Ye et al. (2014)). To test the reliability and precision of the predictive performance, Laio and Tamea (2007) recommended using PQQ plots (Thyer et al., 2009). In the PQQ plot context, if the predictive distribution and observed data are consistent, the corresponding p-value distribution should be uniformly distributed over the whole interval [0,1]. In the present study, the Kolmogorov-Smirnov test (K-S) was used to check this uniformity.

4.3 Case study

4.3.1 The catchment

The applicability and usefulness of the ABC post-processor were assessed over the upper Oria River catchment Figure 17. The River Oria is located in the Basque Country in North of Spain and drains into the Bay of Biscay. This illustrative catchment has an area of 73 km^2 and an elevation range between 180.46 and 1411.57 m.a.s.l. (see the topography in Figure 17).

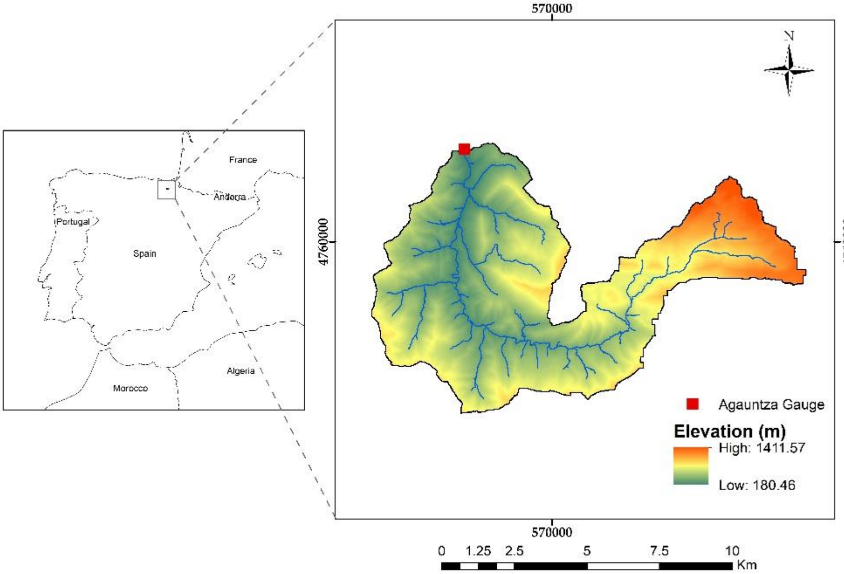

Figure 17. Location and topography of the upper River Oria catchment in the Bay of Biscay (Spain).

The climate of this area is oceanic, with the mean annual discharge, reference evapotranspiration and precipitation for the period 1987–2000 of 765, 733.4 and 1498 mm, respectively. Monthly variability is shown in Figure 18. Aridity and the run-off index are 0.489 and 0.511, respectively. Land cover includes grass (32%), natural forest (20.5%) and forestry plantations (38.6%). Soils are mostly cambisols (38.5%) and luvisols (47.0%). For reviews of the study area, see (Francés et al., 2007).

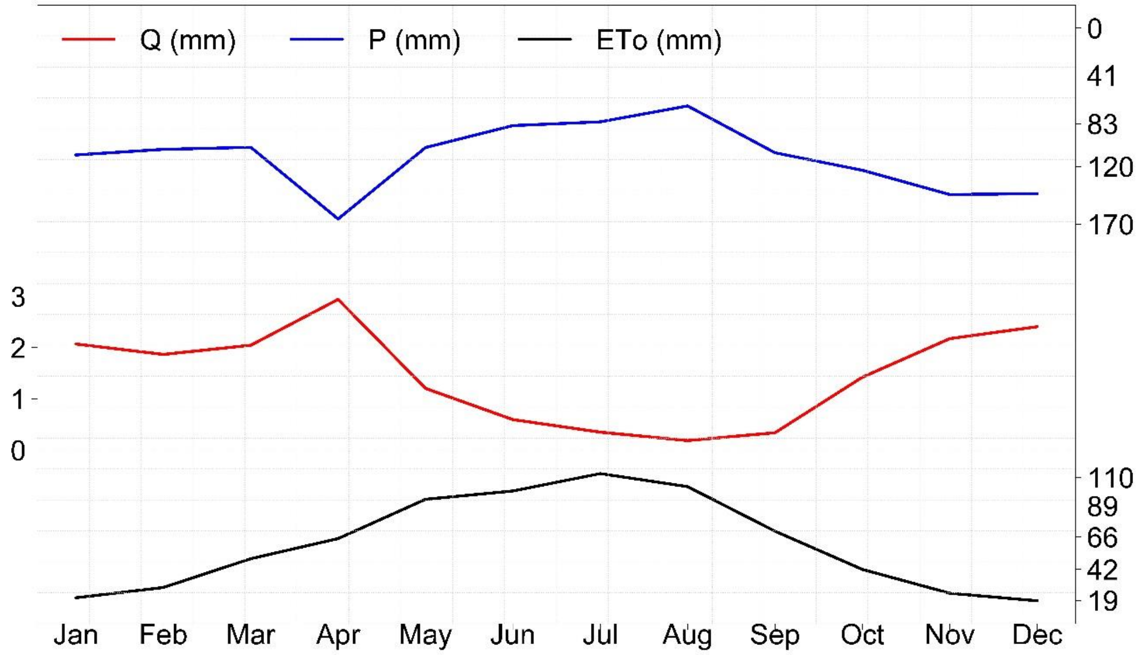

Figure 18. Mean monthly values (1987–2000) for the upper River Oria catchment of precipitation (P), reference evapotranspiration (ETo) and streamflow (Q).

4.3.2 Observational data and hydrological model

Daily precipitation, temperature and reference evapotranspiration data were retrieved from the State Meteorological Agency of Spanish Government - AEMET (http://www.aemet.es). Potential evapotranspiration was estimated by the Hargreaves method (Hargreaves and Samani, 1982) from maximum and minimum temperatures, but was calibrated in the nearest meteorological station. Finally, the daily observed streamflows were collected from the Provincial Guipuzcoa Government. The hydrological simulations for this study were grounded on a conceptually based and spatially distributed model called TETIS, which simulated the main hydrological process. In TETIS, each grid cell represents a tank model with six tanks connected to one another (Francés et al., 2007). TETIS was developed by our Research Group (http://lluvia.dihma.upv.es) over the last 20 years, with worthy results in different climate and catchments size scenarios (Bussi et al., 2014; Francés et al., 2007; Rodriguez-Lloveras et al., 2015; Ruiz-Pérez et al., 2016; Ruiz-Villanueva et al., 2015; Vélez et al., 2009). The model was calibrated against discharge using the automatic optimisation algorithm SCE-UA (Duan et al., 1994) in daily time steps for the 1990–2000 period, and was validated for the 1987–1990 period. The objective function was the Nash Sutcliffe Efficiency (NSE) Index. For further details on model implementation in this case study, refer to Francés et al. (2007) and Vélez et al. (2009). For this chapter, the simulated daily streamflows from TETIS were aggregated to the monthly data at the C2Z1 Agauntza gauge (the catchment outlet in this case study).

4.3.3 Climate change data

The climate change projections were obtained from the Spanish National Meteorological Agency - AEMET (http://www.aemet.es), which provides the regionalised projections of 14 climate models (both GCM and RCM) using the statistical technique of analogues (Lorenz, 1969). These projections were obtained from the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change - IPCC (Magrin et al., 2014) and the CORDEX Project (Colmet-Daage et al., 2018). The historical baseline period was 1987–2000. The daily precipitation, and the minimum and maximum temperatures for the baseline period were used. To remove the bias of the climate model simulations, the quantile mapping method (Gudmundsson et al., 2012) was employed. Next all the models simulated the long-term daily streamflow for the 1987–2000 period using 14 downscaled climate data. These daily data were also aggregated to the monthly time series. Finally, the ABC post-processor was applied to post-process the projected streamflows and to assess their conditional predictive uncertainty during the baseline period (1987–2000). Detailed information about the 14 climate models is listed in Table 5.

Table 5. Information about the 14 selected climate models.

Code	Model Name	Modelling group
1	**ACCESS1**	Commonwealth Scientific and Industrial Research Organization/ Bureau of Meteorology, Australia
2	**bcc_csm1**	Beijing Climate Center, China
3	**BNU_ESM**	Beijing Normal University, China
4	**CMCC_CESM**	Ceatre Europeen de Recherche et Formation Avabcees en Calcul Scientifique
5	**CMCC_CM**	
6	**CNRM_CM5**	
7	**GFDL_ESM2G**	Geophysical Fluid Dynamics Laboratory, USA
8	**GFDL_ESM2M**	Geophysical Fluid Dynamics Laboratory, USA
9	**inmcm4**	Meteorological Research Institute
10	**IPSL_CM5A_MR**	Institute Pierre-Simon Laplace, France
11	**MIROC_ESM**	National Institute for Environmental Studies
12	**MIROC5**	
13	**MPI_ESM_MR**	Max Planck Institute for Meteorology
14	**MRI_CGCM3**	Meteorological Research Institute

4.4 Results

4.4.1 Hydrological model evaluation

Figure 19 shows the daily TETIS performance for the calibration period (1987 – 2000) in the case study with a monthly aggregation of streamflows. An accurate agreement in timing between the observed and simulated monthly streamflows is noticed. The TETIS parameters were calibrated and validated, to produce a Nash-Sutcliffe Efficiency (NSE) and a Kling-Gupta Efficiency (KGE) at monthly discretisation, equal 0.93 and 0.87 for the calibration period, and to 0.88 and 0.86 for the validation period. These performances are considered “good” on the basis of the performance measures proposed by Martinez and Gupta (2010). The latter can be confirmed looking at the scatterplot of monthly streamflows (Figure 19, right), with the majority of points closely following the bisector. Compared to the calibration results, the NSE and KGE indices are slightly degraded during the validation period, which is as expected for adequate model calibration.

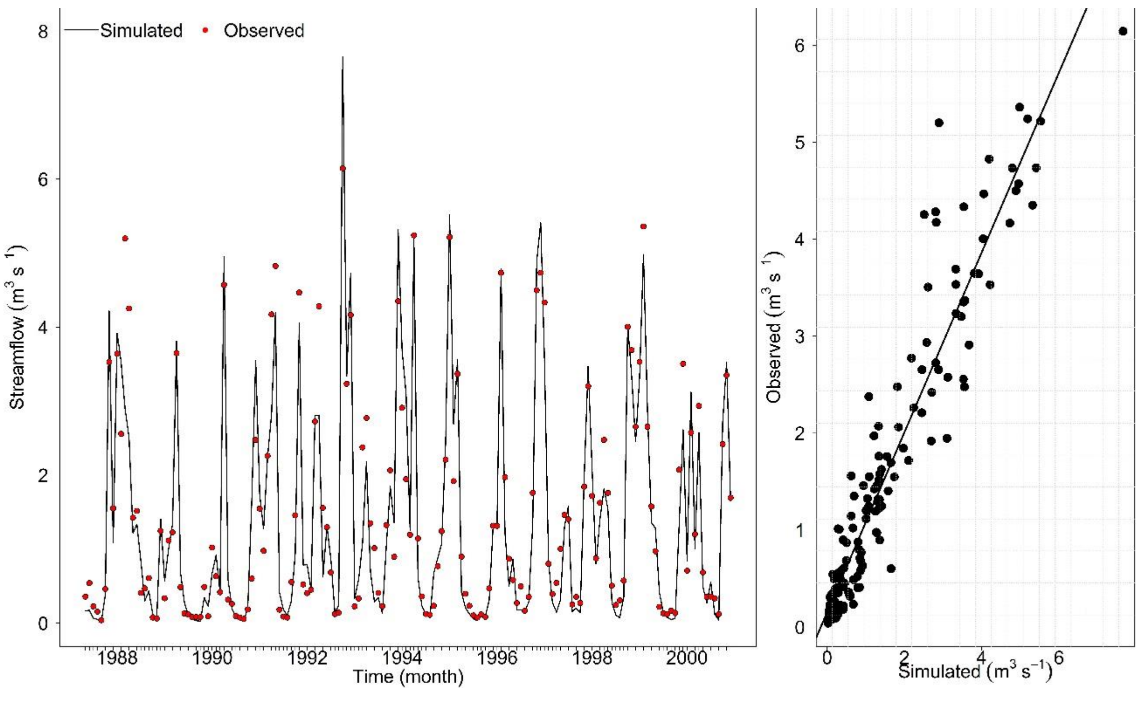

Figure 19. Left panel: monthly time series for the baseline period of the hydrological predictions (solid line) and observations (red dots) for the case study. Right panel: scatter plot of simulated *versus* observed streamflows.

According to the above results, the hydrological parameters are reasonably acceptable and can be further conducted with future hydrological simulations driven by projected meteorological data from climate models.

4.4.2 The general streamflow predictive performance of the climate models

The TETIS simulations for the baseline period (1987–2000), forced with 14 climate model outputs, were compared with the observed streamflows in the case study outlet. Figure 20 shows the observed and simulated mean monthly streamflows based on 14 climate models. In Figure 20, the label called "Reference" indicates the hydrological simulations forced with the observed hydrometeorological data, while "Mean ensemble" indicates the mean of the 14 climate models during the baseline period. In general, the 14 climate models fairly represented the mean seasonal streamflow patterns of the case study. However, the data in Figure 20 show that most of the climate models led to considerably overestimated streamflows in winter and spring, and to slightly underestimated streamflows in summer and autumn.

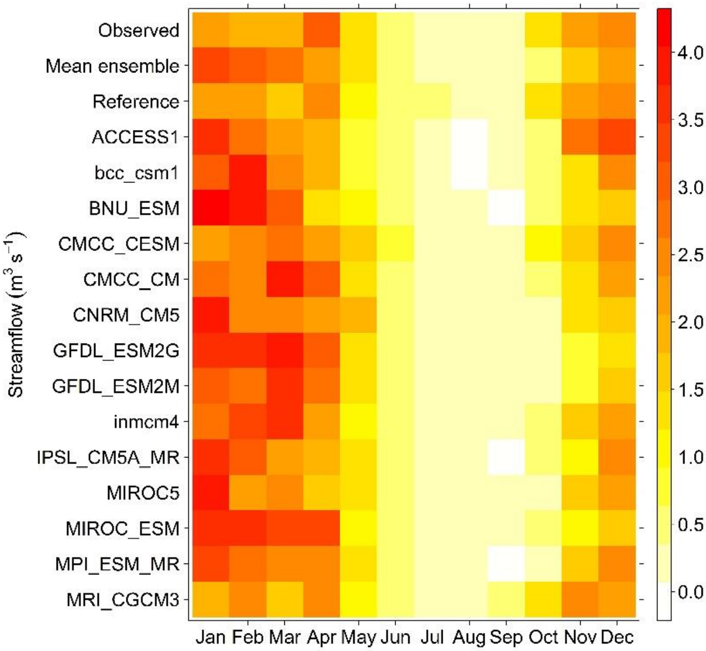

Figure 20. The mean monthly streamflows during the baseline period (1987–2000) for the case study by 14 climate models compared with the observed and ensemble means.

These results suggest that climate models present a seasonal bias. In contrast, CMCC_CESM clearly outperformed the other climate models and mimicked the mean seasonality cycles that, albeit not perfect, came closer to the observed streamflow (see also Figure 21). Most climate models led to a considerable overestimated streamflow during high flows and to an underestimated streamflow during low flows (Figure 21). This result implies that few climate models that projected streamflows are not directly applicable for long-term water management and adaptation planning.

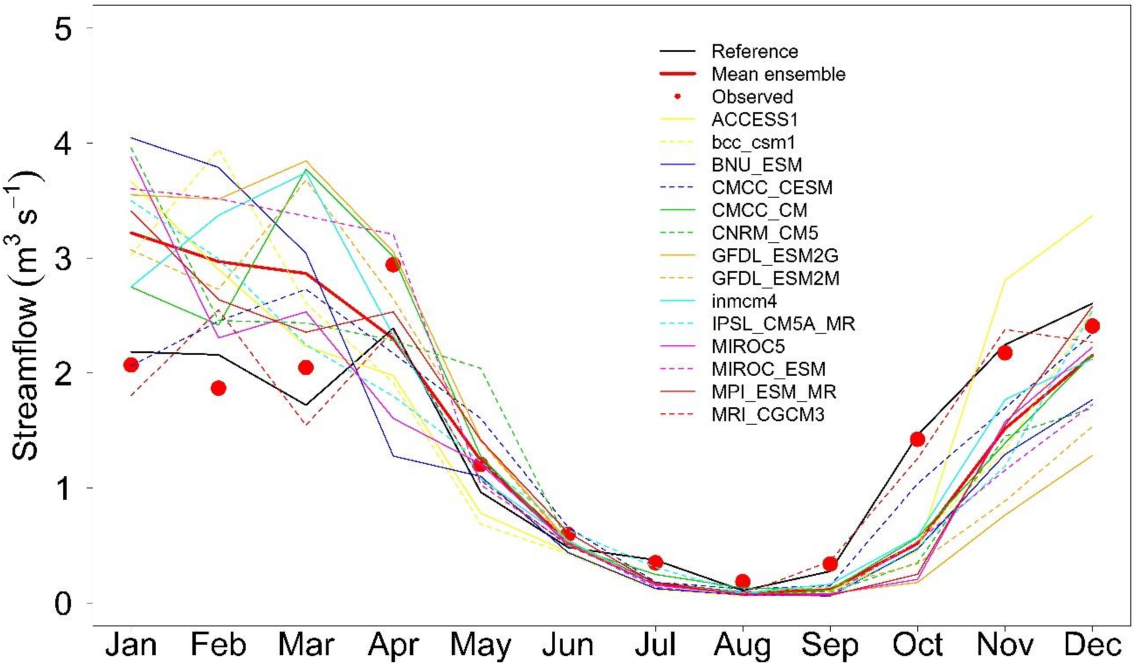

Figure 21. The mean monthly streamflow hydrograph during the baseline period (1987–2000) for the case study by 14 climate models compared with the observed and ensemble means.

Figure 22 comparing the NSE performance of the 14 climate models to the post-processed streamflow (the predictive uncertainty median). The results generally showed that the NSE performance index for the post-processed streamflow were mostly better than that for the 14 climate models streamflow (without post-processing). This result indicates that the ABC post-processor implicitly reduces bias. The ABC post-processor usually increases the NSE index. However if the climate model initially displays high performance (without post-processing); then ABC post-processor performance is not significant. For example, the CMCC_CESM model, which the NSE barely increased by 0.4%. In contrast, NSE for MIROC_ESM increased by as much as 77.2%. The KGE index showed overall similar behaviour, but is not presented herein for the sake of brevity.

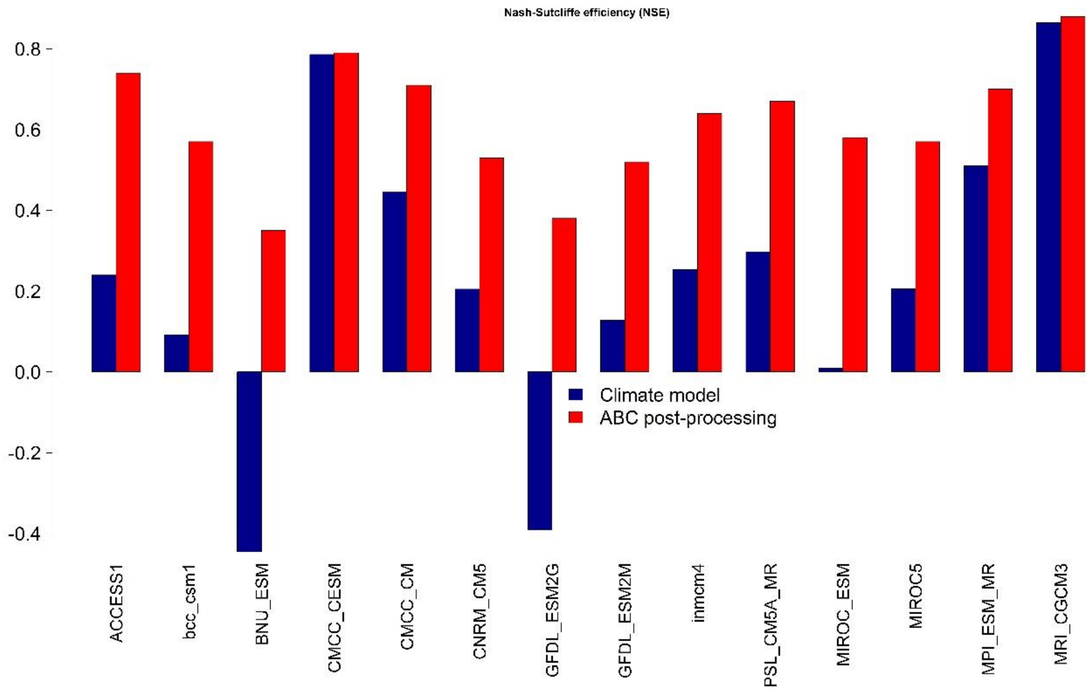

Figure 22. The NSE performance index of the 14 climate models and the post-processed streamflow simulations for the baseline period (1987–2000).

After we showed the poor performance of the 14 climate models (Figure 20, Figure 21 and Figure 22), the ABC post-processor was applied to improve the predictive ability of climate models, and its predictive performance was assessed using the verification metrics explained in the Methodology (see Section 4.2.3). Table 6 shows the overall performance results of all the 14 climate models and the "Reference", e.g., the hydrological simulations forced with the observed hydrometeorological data. To compute the deterministic metrics (NSE and KGE), predictive distribution median was used. The results indicated that the variation between the climate models was wide in terms of the NSE, KGE and CR (95%) metrics. Specifically, the NSE ranged from 0.35 (BNU_ESM) to 0.88 (MRI_CGCM3) with 53% of variability, the KGE ranged from 0.69 (BNU_ESM) to 0.89 (MRI_CGCM3) with 20% of variability, and the CR ranged from 50%

(GFDL_ESM2G and GFDL_ESM2M) to 91.7% (CMCC_CESM) with 41.7% of variability.

Table 6. The deterministic and probabilistic performance metrics of the simulated long-term mean streamflows based on the 14 climate models for the case study.

Model	NSE	KGE	CR (95%)	Precision	K-S test	Reliability
Reference	0.93	0.87	100.00	5.49	0.30	0.94
ACCESS1	0.74	0.87	75.00	5.37	0.42	0.95
bcc_csm1	0.57	0.78	58.30	5.40	0.43	0.96
BNU_ESM	0.35	0.69	58.30	5.30	0.24	0.96
CMCC_CESM	0.79	0.87	91.70	5.56	0.21	0.96
CMCC_CM	0.71	0.85	75.00	5.43	0.09	0.96
CNRM_CM5	0.53	0.79	58.30	5.37	0.17	0.96
GFDL_ESM2G	0.38	0.70	50.00	5.39	0.02	0.96
GFDL_ESM2M	0.52	0.76	50.00	5.41	0.04	0.96
inmcm4	0.64	0.82	66.70	5.52	0.18	0.96
IPSL_CM5A_MR	0.67	0.83	58.30	5.45	0.03	0.96
MIROC_ESM	0.58	0.77	58.30	5.30	0.33	0.97
MIROC5	0.57	0.79	75.00	5.32	0.10	0.10
MPI_ESM_MR	0.70	0.85	66.70	5.37	0.17	0.96
MRI_CGCM3	0.88	0.89	83.30	5.44	0.89	0.94

In general, MRI_CGCM3 outperformed the other climate models in deterministic metrics terms (NSE and KGE) and CMCC_CESM did so in probabilistic metric terms (CR). If the predictive distribution and observed data are consistent, the corresponding p-value distribution should be uniformly distributed over the interval [0, 1]. Table 6 indicates that only IPSL_CM5A_MR, GFDL_ESM2G and GFDL_ESM2M did not pass the K-S test to check the uniformity distribution (p-value < 0.05). According to the precision and reliability metrics, no crucial differences were found among the 14 climate models (Table 6). The conditional predictive PQQ-plots of the ensemble of climate simulations are shown in Figure 23. Four climate models (MRI_CGCM3, CMCC_CESM, ACCESS1 and bcc_csm1) gave a correct predictive uncertainty estimation, while the other climate models underestimated the predictive uncertainty using the ABC post-processor. It is clear from the Figure 23 that only a few curves closely follow the bisector lines. This indicates that the conditional predictive uncertainty of the climate streamflow simulations is reliable. In Figure 23 the panel called “Reference” denotes the hydrological simulations forced with observed inputs.

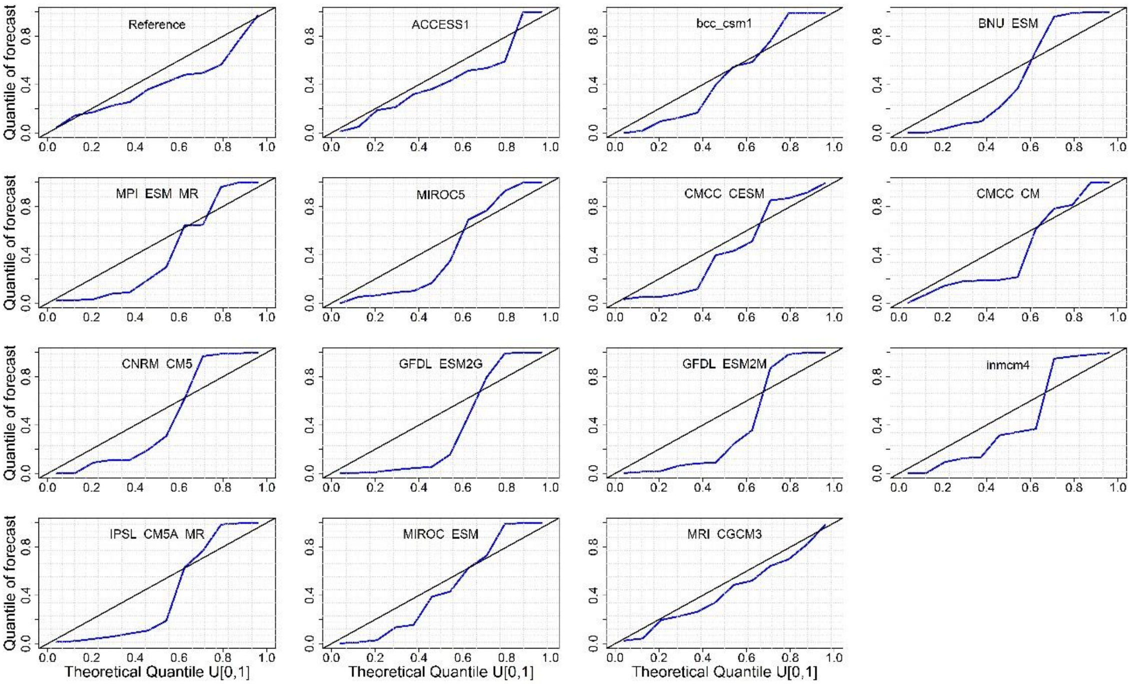

Figure 23. The Conditional predictive PQQ-plot for the baseline period (1987–2000) in the case study for the hydrological model forced by the 14 climate models and the reference.

To maintain simplicity and clarity, and to extend uncertainty analysis results, detailed results are presented for three selected climate models. Therefore, three class of climate model results are presented in the following subsections, which range from the categories “good” (CMCC_CESM) to “mediocre” (ACCESS1) and “poor” (GFDL_ESM2G) according to the predictive performances for the baseline period.

4.4.2.1 Predictive performance of CMCC_CESM (good category)

In this subsection, the conditional predictive uncertainty of CMCC_CESM was assessed for the baseline period. Figure 24 displays the observed and simulated long-term mean monthly streamflows (1987–2000) based on the CMCC_CESM climate model and the conditional predictive uncertainty for CMCC_CESM.

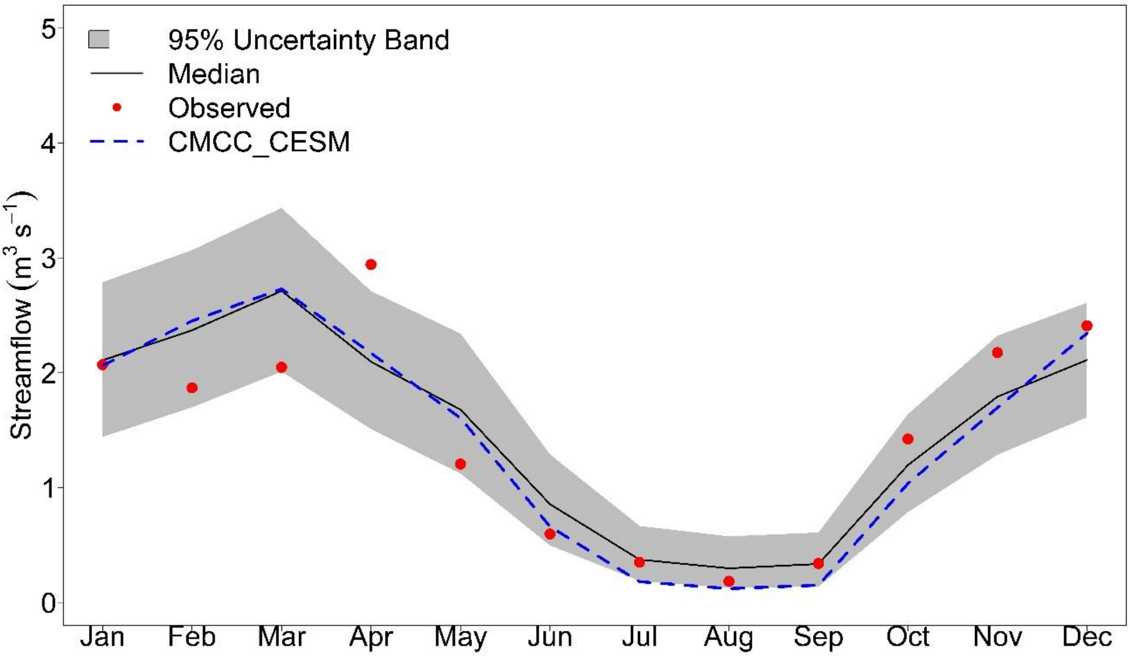

Figure 24. The observed and simulated long-term mean monthly streamflows (1987–2000) based on the CMCC_CESM climate model (without post-processing, blue line). The conditional predictive uncertainty for CMCC_CESM (with post-processing, black line).

The ABC post-processor produced narrow predictive uncertainty bounds with a containing ratio (CR) of 91.7% (Table 6). We should bear in mind that perfect predictive distribution would expect that the CR would come close to the assumed 95% prediction level. It is evident from the results that the ABC post-processor rectifies the over-/under estimated streamflow of the the CMCC_CESM climate model (without post-processing). Moreover, the predictive uncertainty median supplies an adequate description of the observed streamflows (black line). This result indicates that the ABC post-processor implicitly reduces the bias. Although the CMCC_CESM performance is suitable (Figure 22), the use of the ABC post-processor to provide the conditional predictive uncertainty, which is essential for decision making. When the ABC post-processor was used, NSE and

KGE barely increased by 0.4% and by 1.7%, respectively, which suggests that the high performance of CMCC_CESM limits the ABC post-processor performance in terms of its bias correction.

4.4.2.2 Predictive performance of ACCESS1 (mediocre category)

Figure 25 shows the long-term mean monthly streamflow time series and the conditional predictive uncertainty for ACCESS1 (yellow line). As in the CMCC_CESM model (Subsection 4.4.2.1), the ABC post-processor significantly improved the predictive performance compared to the original ACCESS1 (without post-processing) (Figure 22). NSE increased by 50% and KGE rose by 40%. Conversely in probabilistic metrics terms, the CR of ACCES1 was lower than CMCC_CESM (Table 6). These results suggest that the quality of climate models influences the ABC post-processor's predictive performance.

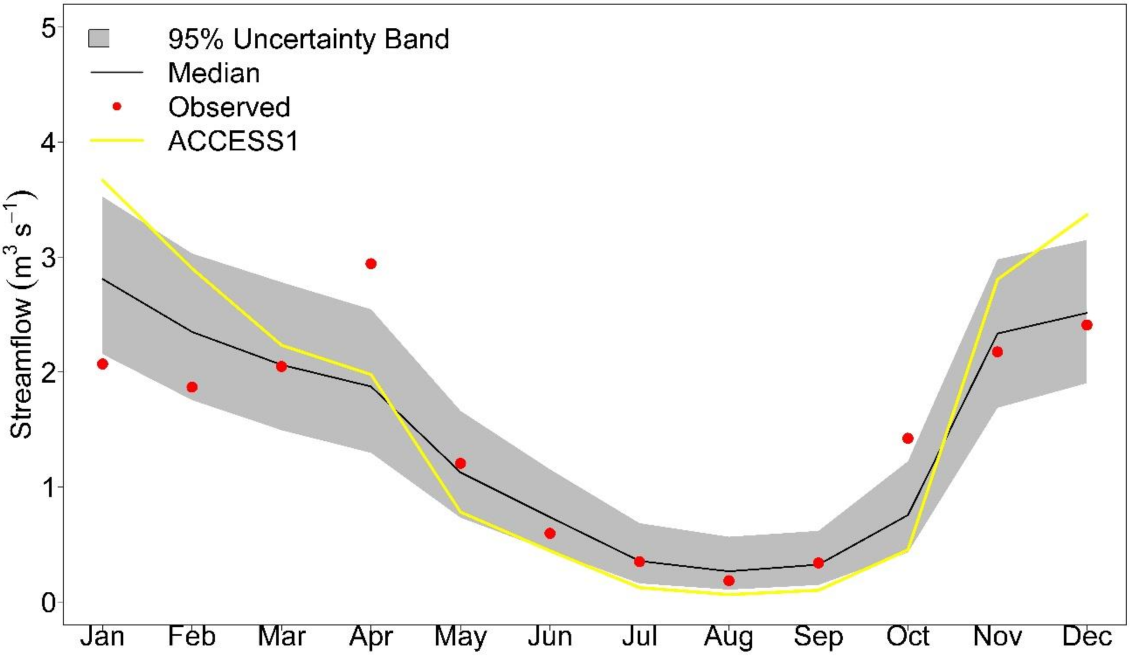

Figure 25. The observed and simulated long-term mean monthly streamflows (1987–2000) based on the ACCESS1 climate model (without post-processing,

yellow line). The conditional predictive uncertainty for ACCESS1 (with post-processing, black line).

4.4.2.3 Predictive performance of GFDL_ESM2G (poor category)

Figure 26 displays the long-term mean monthly streamflow time series and the conditional predictive uncertainty for GFDL_ESM2G (orange line). As shown in Figure 26, the ABC post-processor also made vague predictions, with a poorly predictive performance and uncertainty quantification. The CR performed the worst with 50% (Table 2) because GFDL_ESM2G deviated considerably from the observed streamflows. Consequently, the previous results (Subsection 4.4.2.2) confirmed that the climate model's quality influenced the uncertainty quantification and predictive performance of the ABC post-processor. In contrast, NSE increased by as much as 77.2% and KGE rose by 40.5% (Figure 22).

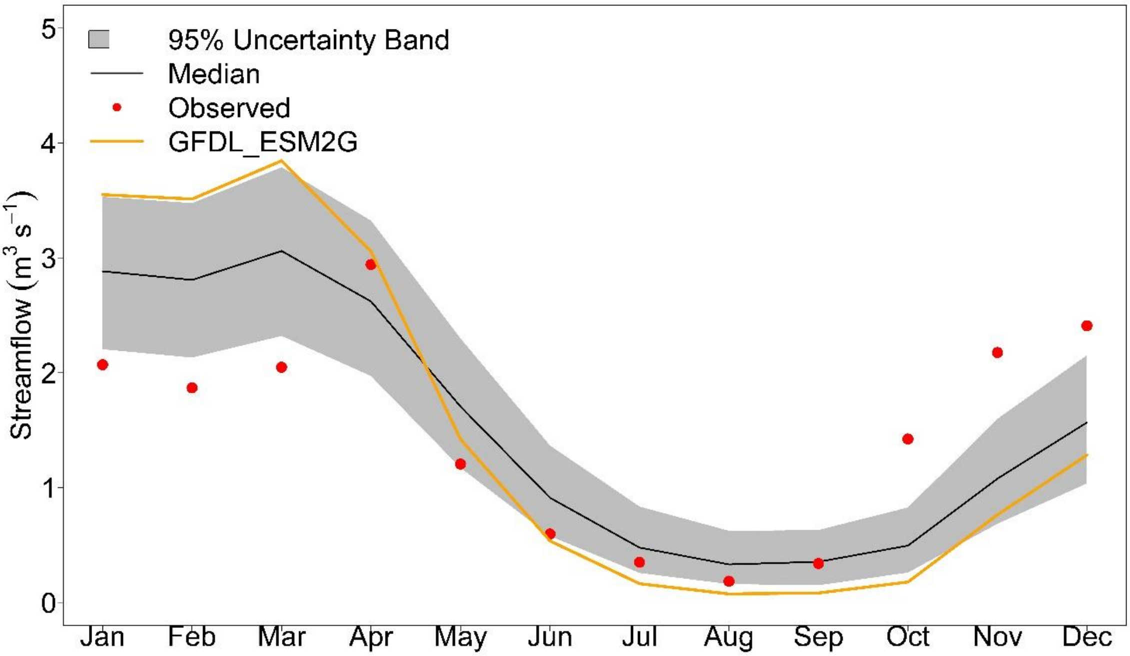

Figure 26. The observed and simulated long-term mean monthly streamflows (1987–2000) based on the GFDL_ESM2G climate model (without post-processing, orange line). The conditional predictive uncertainty for GFDL_ESM2G (with post-processing, black line).

4.5 Discussion and Conclusions

We found that the ABC post-processor approach offered more reliable projections than 14 climate models (without post-processing) (Figure 22). Such a result must have been due to the post-processing method working directly to improve the errors in the model outputs (Ye et al., 2014). This is not a surprising finding because the post-processing method has proven its capacity to improve predictions and to eliminate systematic errors, which are inherent to hydrological and environmental models (Bogner et al., 2016; Woldemeskel et al., 2018; Ye et al., 2014).

The traditional climate change impact studies assume that a range of GCM-RCM chains properly represents the uncertainty of climate change projections. However, climate change projections are not in synchrony with observations (Maraun, 2016), so climate change projections model streamflow statistics instead of time series. Unfortunately, a little information is available on the uncertainty of streamflow statistics. To bridge this gap, this study introduces a coherent univariate method, called ABC post-processor, to assess the conditional predictive uncertainty of streamflow statistics obtained from climate change projections that merge statistical post-processing, approximate Bayesian computation and streamflow statistics.

Unlike the results for the predictive accuracy, the reliability of the uncertainty band depended on climate models' quality. Essentially, the ABC post-processor assumes that climate models correlate with the streamflow statistics of the observations. However, this assumption is not valid for climate models that perform poorly during the baseline period (e.g. GFDL_ESM2G, BNU_ESM, MIROC_ESM). This is because the ABC post-processor and many other post-processors are pure statistical models and cannot correct the fundamental problems of mechanistic climate models (Maraun, 2016). In fact the ABC post-processor produced more realistic uncertainty bands than the 14 climate models (without post-processing) when the climate models performed well during the baseline period.

In our study, the 14 climate models (GCM/RCM) showed a systematic seasonal bias, with overestimation noticed for high flows and underestimation for low flows. One possible explanation for this is that precipitation strongly depends on the representation of the still unresolved convective processes (Benedict et al., 2019). In addition, spatio-temporal and multi-variable aspects are frequently misrepresented by climate models (Maraun, 2016). This finding falls in line with the results of Dang et al., (2017) and Gao et al., (2019). As expected, the 14 climate models (without post-processing) were more biased than the post-processed streamflows (Figure 22). This finding coincides with the outcomes of Ahn and Kim (2019), who stated that multi-model climate ensembles were a biased information approach. It was herein proved that the ABC post-processor was able to improve climate change projected streamflows by reducing bias (Figure 22) and by producing realistic uncertainty bands (Table 6).

Based on our results, the 14 climate models showed large differences in the magnitudes of the projected streamflows (Table 6). This spread can be associated with the selection and the parameterisation of the climate models (GCM/RCM). Inevitably, removing this bias is necessary to derive robust impact studies. Additionally, the bias of the 14 climate models suggests that the unweighted ensemble approach, which is an equally probable climate model, would be questionable for monthly streamflows. This result was confirmed by Knutti et al. (2019), who stated that the main problem of the unweighted ensemble approach, which is also called the model democracy or “one model one vote”, lies in treating all the models as being independent and equally plausible. Biondi and Todini (2018) suggested that assumptions of equal probability were violated as a consequence of many sources of uncertainty (inputs, outputs, epistemic, initial conditions, and so forth).

This study was limited to the post-processing method for a univariate framework, which is one predictor (climate change model’s output). Nevertheless, the ABC post-processor is very flexible and can be used for more complex models. Climate change impact assessments involves many important sources of uncertainty: hydrological model, emission scenario, GCM downscaling method, bias correction method and internal climate variability. However, the present study focused only on the conditional predictive uncertainty of the streamflows obtained from climate models (GCM/RCM), without including the uncertainty of emissions or RCP scenarios.

Future researches will extent the ABC post-processor to handle multiple climate models (multivariate framework) to post-process multi-model ensembles. While our research focused on the predictive uncertainty of projected streamflows, future works will investigate the effect of pre-processing inputs (i.e. precipitation and temperature) and post-processing streamflows to identify the main source of predictive uncertainty, such as Lucatero et al., (2018) obtained for seasonal forecasts. Forthcoming studies will also implement the ABC post-processor in other catchments with different hydrological conditions to be able to draw more generalised conclusions. Note that the ABC post-processing method could help to transfer predictive uncertainty from gauged to ungauged catchments as in (Smith et al., 2014), namely transfer modelling error characteristics (Bourgin et al., 2015). At the same time, the ABC post-processor could be adapted as a method to constrain ensembles: see for instance, Yadav et al. (Yadav et al., 2007), Kapangaziwiri et al. (2012), Padrón et al. (2019), and (Zhang et al., 2008).

Finally, the present research does not consider that the 14 climate models are useless for assessing predictive uncertainty of streamflow statistics. Rather they are valuable as a first approximation, but definitely do not suffice. As pointed out by Maraun (2010), a model consensus did not imply reliability as deficiencies are common to all climate models. Knutti et al. (2019) also discussed how multi-model spread is merely a range of across models and cannot be understood as an uncertainty analysis. Similarly, Nearing and Gupta (2018) advocated that the multi-model approach is a kind of sensitivity analysis, that is not interpreted as an uncertainty analysis. Biondi and Todini (2018) argued that by using only the ensemble address method with the input values to estimate uncertainty, it does not correctly quantify the predictive uncertainty, and underestimated predictions can be expected. In summary, we recommended using the ABC post-processor to complement the traditional multi-model method to assess the conditional predictive uncertainty of climate change projections. In other words, the ABC post-processor can improve climate change projected streamflows by reducing bias and producing more realistic uncertainty bands but only for individual members of the ensemble. This framework can be used to provide essential water and environmental resources management information, to plan and operate water resources systems and to support robust decisions about adaptation plans.

CHAPTER 5. Conclusions

5.1 Concluding remarks

The principal focus of this thesis is on improved hydrological post-processing methods for assessing the conditional predictive uncertainty of monthly streamflows. In particular, we address two main issues i) the heteroscedasticity problem and ii) the intractable likelihood problem.

In chapter 2, we deal with the heteroscedasticity of variance, which means the prediction uncertainty increases with the magnitude of prediction variables (Coccia and Todini, 2011). To overcome this issue, we develop the GMM post-processor, which is based on the Bayesian joint probability modelling approach and the Gaussian mixture models. The proposed post-processor is found to be the best suited for estimating the conditional predictive uncertainty of monthly streamflows, especially for dry catchments. The proposed post-processor delivers the sharpest predictive uncertainty without sacrificing reliability across 12 MOPEX catchments with different hydroclimatic and physical conditions. Although GMM post-processor is not a parsimonious approach, we recommend its use, especially in dry catchments.

In chapter 3, we develop a second post-processor to address the intractable likelihood problem, which means the likelihood function is unavailable in closed form or by numerical derivation (Robert, 2016), called ABC post-processor, based on summary statistics and a free-likelihood approach to estimate the conditional predictive uncertainty of monthly streamflows. The proposed post-processing method exchanges the requirement of computing the likelihood function by the use of some sufficient summary statistics and synthetic datasets. We proved that the conditional predictive distribution is qualitatively similar produced by the exact predictive (MCMC post-processor) or the approximate predictive (ABC post-processor). This finding is significant because dealing with scarce information is a common condition in hydrological studies. Therefore, we recommend applying the ABC post-processor when it is cumbersome to calculate the likelihood function.

Finally, in chapter 4, we applied the ABC post-processor, to assess the conditional predictive uncertainty of streamflow statistics obtained from climate change projections that merge statistical post-processing, approximate Bayesian computation and streamflow statistics, such as a particular case of intractable likelihood problem. We analysed the climate change projections (AR5 - IPCC) of the monthly streamflows in the upper Oria catchment (Spain). We found that the ABC post-processor approach: 1) offered more reliable projections than 14 climate models (without post-processing) 2) for the best climate models during the baseline period produced more realistic uncertainty bands than the classical multi-model ensemble approach. So, we recommend using the ABC post-processor to complement the traditional multi-model method to assess the conditional predictive uncertainty of climate change projections.

Overall, this thesis is established upon the foundation that hydrological predictions can be more valuable when expressed in probabilistic terms of probability distribution function (PDF) or terms of predictive uncertainty bands. So, the main contribution of this research is the simple integration of process-based models (deterministic models) and statistical post-processing for probabilistic hydrological modelling. In this framework, statistical post-processing methods are applied to convert the point predictions provided by hydrological models to probabilistic forecasts. This contribution develops more reliable hydrological predictions and adequate representation of predictive uncertainty. Besides, it is essential to mention that all statistical post-processors implemented in this thesis are available in open source; therefore, their reproducibility is entirely guaranteed. The analyses and visualisations have been performed in R Programming Language.

5.2 Future research lines

There are several opportunities for further improvements in the proposed hydrological post-processing methods used in this thesis. The two proposed post-processing methods are a univariate approach, which handles only one hydrological model outputs (predictor). So, further work is needed to extend the proposed post-processors to handle multiple predictors, which is a multivariate model approach. Besides, future comparison studies are also necessary to evaluate the proposed post-processors with different hydroclimatic variables (i.e., soil moisture, water level, precipitation, temperature, etc.) and different temporal scale (i.e., hourly, daily, weekly, etc.). Also, the post-processing of extreme events, zero flows and missing data can be a future challenge.

The post-processing methods used in this thesis apply the Normal quantile transformation (NQT) to move the variables to the Normal space. Nevertheless, Brown and Seo (2013) and Madadgar and Moradkhani (2014) argued that any transformation might affect the accuracy of the estimated predictive uncertainty. Thus, future research should develop methods that relax the need for transformation of variables. Moreover, we assumed stationary in the proposed post-processors, but this assumption sometimes is not adequate in hydrology (Li et al., 2017). Consequently, new post-processing approaches are needed to handle non-stationary conditions.

Finally, future research is needed to implement ABC post-processor to estimate the predictive uncertainty of ungauged catchments. This line of work is based on the fact that transferring streamflow statistics is easier than transferring the complete streamflow time series.

6 List of scientific publication

6.1 Peer-reviewed journal publications

Romero-Cuellar, J., Abbruzzo, A., Adelfio, G., Francés, F (2019) Hydrological post-processing based on approximate Bayesian computation (ABC), Stoch Environ Res Risk Assess, 33, 1361-1373, DOI: 10.1007/s00477-019-01694-y.

Romero-Cuellar, J., Francés, F. Estimating the uncertainty of monthly streamflows from climate projections using a post-processing approach. Journal of Hydrology, peer review.

Romero-Cuellar, J., Gastulo-Tapia CJ., Hernández-López, MR., Francés, F. Assessing post-processing approaches for monthly streamflow in 12 MOPEX catchments. Water Resources Research, peer review.

Romero-Cuellar, J., Buitrago-Vargas, A., Quintero-Ruiz, T., Francés, F (2018) Modelling the potential impacts of climate change on the hydrology of the Aipe river basin in Huila, Colombia, Ribagua, 5:1, 63-78, DOI: 10.1080/23863781.2018.1454574.

6.2 Conference contributions and publications in proceedings

Hernández-López, M.R., **Romero-Cuellar, J**., Múnera-Estrada, J.C., Coccia, G., Francés, F. (2017). Performance of two predictive uncertainty estimation approaches for conceptual Rainfall-Runoff Model: Bayesian Joint Inference and Hydrologic Uncertainty Post-processing. Geophysical Research Abstracts. Vol. 19, EGU2017-14782.

Romero-Cuellar, J., Adelfio, G., Francés, F. (2018). Estimating predictive hydrological uncertainty by dressing a probabilistic post-processing approach; a comparison with application to a tropical catchment. Geophysical Research Abstracts. Vol. 20, EGU2018-6915.

Romero-Cuellar, J., Abbruzzo, A., Adelfio, G., Francés, F. (2018). Approximate Bayesian computation for forecasting in hydrological models, In proceedings of the 49th Meeting of the Italian Statistical Society (SIS 2018). Palermo, Italy

Romero-Cuellar, J., Abbruzzo, A., Adelfio, G., Francés, F. (2018). Monthly hydrologic post-processor using approximate Bayesian computation, In 2nd VIBASS WORKSHOP. Valencia, Spain.

Romero-Cuellar, J., Francés, F., Gastulo-Tapia CJ., Hernández-López. (2019). A comparison of the performance of different post-processors of monthly flows in some basins of the MOPEX project, In the tenth edition of the STAHY International Workshop. Nanjing, China

7 References

Adler, C.E., Hirsch Hadorn, G., 2014. The IPCC and treatment of uncertainties: topics and sources of dissensus. Wiley Interdiscip. Rev. Clim. Chang. 5, 663–676. https://doi.org/10.1002/wcc.297

Ahn, K.-H., Kim, Y.-O., 2019. Incorporating climate model similarities and hydrologic error models to quantify climate change impacts on future riverine flood risk. J. Hydrol. 570, 118–131. https://doi.org/10.1016/J.JHYDROL.2018.12.061

Allan, J.A., 2003. Virtual Water - the Water, Food, and Trade Nexus. Useful Concept or Misleading Metaphor? Water Int. 28, 106–113. https://doi.org/10.1080/02508060.2003.9724812

Barbetta, S., Coccia, G., Moramarco, T., Todini, E., 2016. Case Study: A Real-Time Flood Forecasting System with Predictive Uncertainty Estimation for the Godavari River, India. Water 8, 463. https://doi.org/10.3390/w8100463

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian Computation in Population Genetics. Genetics 162.

Benedict, I., van Heerwaarden, C.C., Weerts, A.H., Hazeleger, W., 2019. The benefits of spatial resolution increase in global simulations of the hydrological cycle evaluated for the Rhine and Mississippi basins. Hydrol. Earth Syst. Sci. 23, 1779–1800. https://doi.org/10.5194/hess-23-1779-2019

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298. https://doi.org/10.1002/hyp.3360060305

Biondi, D., Todini, E., 2018. Comparing Hydrological Postprocessors Including Ensemble Predictions Into Full Predictive Probability Distribution of Streamflow. Water Resour. Res. 54, 9860–9882. https://doi.org/10.1029/2017WR022432

Blackwell, D., Dubins, L., 1962. Merging of Opinions with Increasing Information. Ann. Math. Stat. 33, 882–886. https://doi.org/10.1214/aoms/1177704456

Blazkova, S., Beven, K., 2009. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. Water Resour. Res. 45, n/a-n/a. https://doi.org/10.1029/2007WR006726

Boberg, F., Christensen, J.H., 2012. Overestimation of Mediterranean summer temperature projections due to model deficiencies. Nat. Clim. Chang. 2, 433–436. https://doi.org/10.1038/nclimate1454

Bogner, K., Liechti, K., Zappa, M., Bogner, K., Liechti, K., Zappa, M., 2016. Post-Processing of Stream Flows in Switzerland with an Emphasis on Low Flows and Floods. Water 8, 115. https://doi.org/10.3390/w8040115

Bogner, K., Pappenberger, F., Cloke, H.L., 2012. Technical Note: The normal quantile transformation and its application in a flood forecasting system. Hydrol. Earth Syst. Sci 16, 1085–1094. https://doi.org/10.5194/hess-16-1085-2012

Bourgin, F., Andréassian, V., Perrin, C., Oudin, L., 2015. Transferring global uncertainty estimates from gauged to ungauged catchments. Hydrol. Earth Syst. Sci. 19, 2535–2546. https://doi.org/10.5194/hess-19-2535-2015

Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. J. R. Stat. Soc. 26, 211–252.

Broderick, C., Murphy, C., Wilby, R.L., Matthews, T., Prudhomme, C., Adamson, M., 2019. Using a Scenario-Neutral Framework to Avoid Potential Maladaptation to Future Flood Risk. Water Resour. Res. 55, 1079–1104. https://doi.org/10.1029/2018WR023623

Brown, J.D., Seo, D.-J., 2013. Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. Hydrol. Process. 27, 83–105. https://doi.org/10.1002/hyp.9263

Brown, J.D., Seo, D.-J., Brown, J.D., Seo, D.-J., 2010. A Nonparametric Postprocessor for Bias Correction of Hydrometeorological and Hydrologic Ensemble Forecasts. J. Hydrometeorol. 11, 642–665. https://doi.org/10.1175/2009JHM1188.1

Buizza, R., 2018. Ensemble Forecasting and the Need for Calibration. Stat. Postprocessing Ensemble Forecast. 15–48. https://doi.org/10.1016/B978-0-12-812372-0.00002-9

Bulygina, N., Gupta, H., 2009. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. Water Resour. Res. 45, W00B13. https://doi.org/10.1029/2007WR006749

Bussi, G., Francés, F., Horel, E., López-Tarazón, J.A., Batalla, R.J., 2014. Modelling the impact of climate change on sediment yield in a highly erodible Mediterranean catchment. J. Soils Sediments 14, 1921–1937. https://doi.org/10.1007/s11368-014-0956-7

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J. Hydrol. 298, 242–266. https://doi.org/10.1016/j.jhydrol.2004.03.042

Chivers, C., 2012. MHadaptive: General Markov Chain Monte Carlo for Bayesian Inference using adaptive Metropolis-Hastings sampling.

Christensen, J., Kjellström, E., Giorgi, F., Lenderink, G., Rummukainen, M., 2010. Weight assignment in regional climate models. Clim. Res. 44, 179–194. https://doi.org/10.3354/cr00916

Clark, M.P., Kavetski, D., 2010. Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. Water Resour. Res. 46, W10510. https://doi.org/10.1029/2009WR008894

Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H. V., Wagener, T., Hay, L.E., 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resour. Res. 44. https://doi.org/10.1029/2007WR006735

Clark, M.P., Wilby, R.L., Gutmann, E.D., Vano, J.A., Gangopadhyay, S., Wood, A.W., Fowler, H.J., Prudhomme, C., Arnold, J.R., Brekke, L.D., 2016. Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. Curr. Clim. Chang. Reports 2, 55–64. https://doi.org/10.1007/s40641-016-0034-x

Coccia, G., Todini, E., 2011. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. Hydrol. Earth Syst. Sci 15, 3253–3274. https://doi.org/10.5194/hess-15-3253-2011

Colmet-Daage, A., Sanchez-Gomez, E., Ricci, S., Llovel, C., Borrell Estupina, V., Quintana-Seguí, P., Llasat, M.C., Servat, E., 2018. Evaluation of uncertainties in mean and extreme precipitation under climate change for northwestern Mediterranean watersheds from high-resolution Med and Euro-CORDEX ensembles. Hydrol. Earth Syst. Sci. 22, 673–687. https://doi.org/10.5194/hess-22-673-2018

Core Team, R., 2013. R: A language and environment for statistical computing.

Cosgrove, W.J., Loucks, D.P., 2015. Water management: Current and future challenges and research directions. Water Resour. Res. 51, 4823–4839. https://doi.org/10.1002/2014WR016869

Csilléry, K., François, O., Blum, M.G.B., 2012. Abc: An R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3, 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x

Dang, Q.T., Laux, P., Kunstmann, H., 2017. Future high- and low-flow estimations for Central Vietnam: a hydro-meteorological modelling chain approach. Hydrol. Sci. J. 62, 1867–1889. https://doi.org/10.1080/02626667.2017.1353696

Deser, C., Phillips, A., Bourdette, V., Teng, H., 2012. Uncertainty in climate change projections: the role of internal variability. Clim. Dyn. 38, 527–546. https://doi.org/10.1007/s00382-010-0977-x

Diaconis, P., Freedman, D., 1986. On the Consistency of Bayes Estimates. Ann. Stat. 14, 1–26. https://doi.org/10.1214/aos/1176349830

Diks, C.G.H., Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stoch. Environ. Res. Risk Assess. 24, 809–820. https://doi.org/10.1007/s00477-010-0378-z

Dooge, J.C.I., 1986. Looking for hydrologic laws. Water Resour. Res. 22, 46S–58S. https://doi.org/10.1029/WR022i09Sp0046S

Drovandi, C.C., Pettitt, A.N., 2011. Likelihood-free Bayesian estimation of multivariate quantile distributions. Comput. Stat. Data Anal. 55, 2541–2556. https://doi.org/10.1016/J.CSDA.2011.03.019

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H.V. V., Gusev, Y.M.M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O.N.N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., Wood, E.F.F., 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2005.07.031

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158, 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Ehlers, L.B., Wani, O., Koch, J., Sonnenborg, T.O., Refsgaard, J.C., 2019. Using a simple post-processor to predict residual uncertainty for multiple hydrological model outputs. Adv. Water Resour. https://doi.org/10.1016/J.ADVWATRES.2019.05.003

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resour. Res. 50, 2350–2375. https://doi.org/10.1002/2013WR014185

Fatichi, S., Ivanov, V.Y., Paschalis, A., Peleg, N., Molnar, P., Rimkus, S., Kim, J., Burlando, P., Caporali, E., 2016. Uncertainty partition challenges the predictability of vital details of climate change. Earth's Futur. 4, 240–251. https://doi.org/10.1002/2015EF000336

Fearnhead, P., Prangle, D., 2012. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. J. R. Stat. Soc. Ser. B (Statistical Methodol. 74, 419–474. https://doi.org/10.1111/j.1467-9868.2011.01010.x

Feng, K., Zhou, J., Liu, Y., Lu, C., He, Z., 2019. Hydrological Uncertainty Processor (HUP) with Estimation of the Marginal Distribution by a Gaussian Mixture Model. Water Resour. Manag. https://doi.org/10.1007/s11269-019-02260-5

Fenicia, F., Kavetski, D., Reichert, P., Albert, C., 2018. Signature-Domain Calibration of Hydrological Models Using Approximate Bayesian Computation: Empirical Analysis of Fundamental Properties. Water Resour. Res. 54, 3958–3987. https://doi.org/10.1002/2017WR021616

Foley, J.A., 2005. Global Consequences of Land Use. Science (80-.). 309, 570–574. https://doi.org/10.1126/science.1111772

Francés, F., Vélez, J.I., Vélez, J.J., 2007. Split-parameter structure for the automatic calibration of distributed hydrological models. J. Hydrol. 332, 226–240.

Franz, K.J., Hogue, T.S., 2011. Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community. Hydrol. Earth Syst. Sci. 15, 3367–3382. https://doi.org/10.5194/hess-15-3367-2011

Frazier, D.T., Maneesoonthorn, W., Martin, G.M., McCabe, B.P.M., 2019. Approximate Bayesian forecasting. Int. J. Forecast. 35, 521–539. https://doi.org/10.1016/J.IJFORECAST.2018.08.003

Gao, C., Liu, L., Ma, D., He, K., Xu, Y.-P., 2019. Assessing responses of hydrological processes to climate change over the southeastern Tibetan Plateau based on resampling of future climate scenarios. Sci. Total Environ. 664, 737–752. https://doi.org/10.1016/J.SCITOTENV.2019.02.013

Gelman, A., Rubin, D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. Stat. Sci. 7, 457–472. https://doi.org/10.1214/ss/1177011136

Gelman, A., Stern, H., Carlin, J., Dunson, D., Vehtari, A., Rubin, D., 2013. Bayesian data analysis. Chapman and Hall/CRC.

Glahn, H.R., Lowry, D.A., Glahn, H.R., Lowry, D.A., 1972. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J. Appl. Meteorol. 11, 1203–1211. https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B (Statistical Methodol. 69, 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., Hero, A.O., 2013. Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. Water Resour. Res. 49, 2253–2273. https://doi.org/10.1002/wrcr.20161

Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. Hydrol. Earth Syst. Sci. 16, 3383–3390. https://doi.org/10.5194/hess-16-3383-2012

Gupta, H. V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/J.JHYDROL.2009.08.003

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., Andréassian, V., 2014. Large-sample hydrology: a need to balance depth with breadth. Hydrol. Earth Syst. Sci. 18, 463–477. https://doi.org/10.5194/hess-18-463-2014

Haario, H., Saksman, E., Tamminen, J., 2001. An Adaptive Metropolis Algorithm. Bernoulli 7, 223. https://doi.org/10.2307/3318737

Hamill, T.M., 2018. Practical Aspects of Statistical Postprocessing, in: Stéphane Vannitsem, Daniel S. Wilks, J.W.M. (Ed.), Statistical Postprocessing of Ensemble Forecasts. Elsevier, pp. 187–217. https://doi.org/10.1016/b978-0-12-812372-0.00007-8

Hargreaves, G.H., Samani, Z.A., 1982. Estimating Potential Evapotranspiration. J. Irrig. Drain. Div. 108, 225–230.

Hernández López, M.R., 2017. Inferencia Bayesiana conjunta de modelos hidrológicos y modelos de error generalizados, para la evaluación de las incertidumbres predictiva y de los parámetros. Universitat Politècnica de València, Valencia (Spain). https://doi.org/10.4995/Thesis/10251/90652

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian Model Averaging: A Tutorial. Stat. Sci. https://doi.org/10.2307/2676803

Honti, M., Stamm, C., Reichert, P., 2013. Integrated uncertainty assessment of discharge predictions with a statistical error model. Water Resour. Res. 49, 4866–4884. https://doi.org/10.1002/wrcr.20374

Hopson, T.M., Wood, A., Weerts, A.H., 2019. Motivation and Overview of Hydrological Ensemble Post-processing, in: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 783–793. https://doi.org/10.1007/978-3-642-39925-1_36

Kapangaziwiri, E., Hughes, D., Wagener, T., 2012. Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa. Hydrol. Sci. J. 57, 1000–1019. https://doi.org/10.1080/02626667.2012.690881

Kattwinkel, M., Reichert, P., 2017. Bayesian parameter inference for individual-based models using a Particle Markov Chain Monte Carlo method. Environ. Model. Softw. 87, 110–119. https://doi.org/10.1016/J.ENVSOFT.2016.11.001

Kavetski, D., 2019. Parameter Estimation and Predictive Uncertainty Quantification in Hydrological Modelling, in: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 481–522. https://doi.org/10.1007/978-3-642-39925-1_25

Kavetski, D., Clark, M.P., 2010. Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. Water Resour. Res. 46, 2009WR008896. https://doi.org/10.1029/2009WR008896

Kavetski, D., Fenicia, F., Clark, M.P., 2011. Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment. Water Resour. Res. 47. https://doi.org/10.1029/2010WR009525

Kavetski, D., Fenicia, F., Reichert, P., Albert, C., 2018. Signature-Domain Calibration of Hydrological Models Using Approximate Bayesian Computation: Theory and Comparison to Existing Applications. Water Resour. Res. 54, 4059–4083. https://doi.org/10.1002/2017WR020528

Kavetski, D., Kuczera, G., Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resour. Res. 42, n/a-n/a. https://doi.org/10.1029/2005WR004368

Kavetski, D., Kuczera, G., Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resour. Res. 42, n/a-n/a. https://doi.org/10.1029/2005WR004376

Khajehei, S., Moradkhani, H., 2017. Towards an improved ensemble precipitation forecast: A probabilistic post-processing approach. J. Hydrol. 546, 476–489. https://doi.org/10.1016/j.jhydrol.2017.01.026

Klein, B., Meissner, D., Kobialka, H.-U., Reggiani, P., 2016. Predictive Uncertainty Estimation of Hydrological Multi-Model Ensembles Using Pair-Copula Construction. Water 8, 125. https://doi.org/10.3390/w8040125

Knutti, R., Baumberger, C., Hirsch Hadorn, G., 2019. Uncertainty Quantification Using Multiple Models—Prospects and Challenges, in: Computer Simulation Validation. Springer, Cham, pp. 835–855. https://doi.org/10.1007/978-3-319-70766-2_34

Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., Kundzewicz, Z.W., 2018. How the performance of hydrological models relates to credibility of projections under climate change. Hydrol. Sci. J. 63, 696–720. https://doi.org/10.1080/02626667.2018.1446214

Krysanova, V., Hattermann, F.F., 2017. Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results. Clim. Change 141, 363–379. https://doi.org/10.1007/s10584-017-1919-y

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour. Res. 35, 2739–2750. https://doi.org/10.1029/1999WR900099

Krzysztofowicz, R., Kelly, K.S., 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. Water Resour. Res. 36, 3265–3277. https://doi.org/10.1029/2000WR900108

Kuczera, G., Kavetski, D., Franks, S., Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. J. Hydrol. 331, 161–177.

Kundzewicz, Z.W., Krysanova, V., Benestad, R.E., Hov, Ø., Piniewski, M., Otto, I.M., 2018. Uncertainty in climate change impacts on water resources.

Environ. Sci. Policy 79, 1–8. https://doi.org/10.1016/J.ENVSCI.2017.10.008

Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrol. Earth Syst. Sci. 11, 1267–1277. https://doi.org/10.5194/hess-11-1267-2007

Latif, M., 2011. Uncertainty in climate change projections. J. Geochemical Explor. 110, 1–7. https://doi.org/10.1016/J.GEXPLO.2010.09.011

Li, B., Liang, Z., He, Y., Hu, L., Zhao, W., Acharya, K., 2017. Comparison of parameter uncertainty analysis techniques for a TOPMODEL application. Stoch. Environ. Res. Risk Assess. 31, 1045–1059. https://doi.org/10.1007/s00477-016-1319-2

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., Di, Z., 2017. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. Wiley Interdiscip. Rev. Water 4, e1246. https://doi.org/10.1002/wat2.1246

Liang, Z., Chang, W., Li, B., 2012. Bayesian flood frequency analysis in the light of model and parameter uncertainties. Stoch. Environ. Res. Risk Assess. 26, 721–730. https://doi.org/10.1007/s00477-011-0552-y

Lindley, D., Smith, F., 1972. Bayes Estimates for the Linear Model. J. R. Stat. Soc. Ser. B 34, 1–41.

Liu, Y., Gupta, H. V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resour. Res. 43, W07401. https://doi.org/10.1029/2006WR005756

Lorenz, E.N., 1969. Atmospheric Predictability as Revealed by Naturally Occurring Analogues. J. Atmos. Sci. 26, 636–646. https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2

Loucks, D.P., van Beek, E., 2017. System Sensitivity and Uncertainty Analysis, in: Water Resource Systems Planning and Management. Springer International Publishing, Cham, pp. 331–374. https://doi.org/10.1007/978-3-319-44234-1_8

Lucatero, D., Madsen, H., Refsgaard, J.C., Kidmose, J., Jensen, K.H., 2018. Seasonal streamflow forecasts in the Ahlergaarde catchment, Denmark: the effect of preprocessing and post-processing on skill and statistical consistency. Hydrol. Earth Syst. Sci. 22, 3601–3617. https://doi.org/10.5194/hess-22-3601-2018

Madadgar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. Water Resour. Res.

50, 9586–9603. https://doi.org/10.1002/2014WR015965

Magrin, G., Marengo, J., Boulanger, J.-P., Buckeirdge, M., Castellanos, E., Poveda, G., Scarano, F., Vicuña, S., 2014. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, in: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge University Press, Cambridge, United Kingdom , pp. 1499–1566.

Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. J. Hydrol. 330, 368–381. https://doi.org/10.1016/J.JHYDROL.2006.04.046

Maraun, D., 2016. Bias Correcting Climate Change Simulations - a Critical Review. Curr. Clim. Chang. Reports 2, 211–220. https://doi.org/10.1007/s40641-016-0050-x

Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L.O., 2017. Towards process-informed bias correction of climate change simulations. Nat. Clim. Chang. 7, 664–773. https://doi.org/10.1038/nclimate3418

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48, RG3003. https://doi.org/10.1029/2009RG000314

Marin, J.-M., Pudlo, P., Robert, C.P., Ryder, R.J., 2012. Approximate Bayesian computational methods. Stat. Comput. 22, 1167–1180. https://doi.org/10.1007/s11222-011-9288-2

Marjoram, P., Molitor, J., Plagnol, V., Tavare, S., 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. U. S. A. 100, 15324–8. https://doi.org/10.1073/pnas.0306899100

Marshall, L., Nott, D., Sharma, A., 2004. A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. Water Resour. Res. 40, W02501. https://doi.org/10.1029/2003WR002378

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

https://doi.org/10.1029/2009WR008294

Massari, C., Maggioni, V., Barbetta, S., Brocca, L., Ciabatta, L., Camici, S., Moramarco, T., Coccia, G., Todini, E., 2019. Complementing near-real time satellite rainfall products with satellite soil moisture-derived rainfall through a Bayesian Inversion approach. J. Hydrol. 573, 341–351. https://doi.org/10.1016/J.JHYDROL.2019.03.038

McInerney, D., Thyer, M., Kavetski, D., Bennett, B., Lerat, J., Gibbs, M., Kuczera, G., 2018. A simplified approach to produce probabilistic hydrological model predictions. Environ. Model. Softw. 109, 306–314. https://doi.org/10.1016/J.ENVSOFT.2018.07.001

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., Kuczera, G., 2017. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. Water Resour. Res. 53, 2199–2239. https://doi.org/10.1002/2016WR019168

McLachlan, G.J., Krishnan, T. (Thriyambakam), 2008. The EM algorithm and extensions, Wiley Series in Probability and Statistics. Wiley-Interscience, Hoboken, NJ, USA. https://doi.org/10.1002/9780470191613

McLachlan, G.J., Peel, D., 2000. Finite Mixture Models, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/0471721182

McMillan, H., Seibert, J., Petersen-Overleir, A., Lang, M., White, P., Snelder, T., Rutherford, K., Krueger, T., Mason, R., Kiang, J., 2017. How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications. Water Resour. Res. 53, 5220–5228. https://doi.org/10.1002/2016WR020328

Mekonnen, M.M., Hoekstra, A.Y., 2016. Four billion people facing severe water scarcity. Sci. Adv. 2, e1500323. https://doi.org/10.1126/sciadv.1500323

Melnykov, V., Maitra, R., 2010. Finite mixture models and model-based clustering. Stat. Surv. 4, 80–116. https://doi.org/10.1214/09-SS053

Mengersen, K.L., Pudlo, P., Robert, C.P., 2013. Bayesian computation via empirical likelihood. Proc. Natl. Acad. Sci. U. S. A. 110, 1321–6. https://doi.org/10.1073/pnas.1208827110

Min, Y.-M., Kryjov, V.N., Park, C.-K., Min, Y.-M., Kryjov, V.N., Park, C.-K., 2009. A Probabilistic Multimodel Ensemble Approach to Seasonal Prediction. Weather Forecast. 24, 812–828. https://doi.org/10.1175/2008WAF2222140.1

Montanari, A., 2011. Uncertainty of Hydrological Predictions, in: Treatise on Water Science. Elsevier, pp. 459–478. https://doi.org/10.1016/B978-0-444-53199-5.00045-2

Montanari, A., Brath, A., 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. Water Resour. Res. 40, W01106. https://doi.org/10.1029/2003WR002540

Montanari, A., Grossi, G., 2008. Estimating the uncertainty of hydrological forecasts: A statistical approach. Water Resour. Res. 44, W00B08. https://doi.org/10.1029/2008WR006897

Montanari, A., Koutsoyiannis, D., 2012. A blueprint for process-based modeling of uncertain hydrological systems. Water Resour. Res. 48, n/a-n/a. https://doi.org/10.1029/2011WR011412

Moradkhani, H., Hsu, K.-L., Gupta, H., Sorooshian, S., 2005a. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour. Res. 41, n/a-n/a. https://doi.org/10.1029/2004WR003604

Moradkhani, H., Nearing, G.S., Abbaszadeh, P., Pathiraja, S., 2019. Fundamentals of Data Assimilation and Theoretical Advances, in: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 675–699. https://doi.org/10.1007/978-3-642-39925-1_30

Moradkhani, H., Sorooshian, S., Gupta, H. V., Houser, P.R., 2005b. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Resour. 28, 135–147. https://doi.org/10.1016/J.ADVWATRES.2004.09.002

Moriasi, D., Arnold, J.G., Liew, M., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. MODEL EVALUATION GUIDELINES FOR SYSTEMATIC QUANTIFICATION OF ACCURACY IN WATERSHED SIMULATIONS.

Muhammad, A., Stadnyk, T., Unduche, F., Coulibaly, P., 2018. Multi-Model Approaches for Improving Seasonal Ensemble Streamflow Prediction Scheme with Various Statistical Post-Processing Techniques in the Canadian Prairie Region. Water 10, 1604. https://doi.org/10.3390/w10111604

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768–772.

https://doi.org/10.1038/nature02771

Najafi, M.R., Moradkhani, H., Jung, I.W., 2011. Assessing the uncertainties of hydrologic model selection in climate change impact studies. Hydrol. Process. 25, 2814–2826. https://doi.org/10.1002/hyp.8043

Nearing, G.S., Gupta, H. V., 2018. Ensembles vs. information theory: supporting science under uncertainty. Front. Earth Sci. 12, 653–660. https://doi.org/10.1007/s11707-018-0709-9

Nott, D.J., Marshall, L., Brown, J., 2012. Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? Water Resour. Res. 48, n/a-n/a. https://doi.org/10.1029/2011WR011128

Padrón, R.S., Gudmundsson, L., Seneviratne, S.I., 2019. Observational Constraints Reduce Likelihood of Extreme Changes in Multidecadal Land Water Availability. Geophys. Res. Lett. 46, 736–744. https://doi.org/10.1029/2018GL080521

Pechlivanidis, I.G., Gupta, H., Bosshard, T., 2018. An Information Theory Approach to Identifying a Representative Subset of Hydro-Climatic Simulations for Impact Modeling Studies. Water Resour. Res. 54, 5422–5435. https://doi.org/10.1029/2017WR022035

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7

Price, L.F., Drovandi, C.C., Lee, A., Nott, D.J., 2018. Bayesian Synthetic Likelihood. J. Comput. Graph. Stat. 27, 1–11. https://doi.org/10.1080/10618600.2017.1302882

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W., 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. 16, 1791–1798. https://doi.org/10.1093/oxfordjournals.molbev.a026091

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Mon. Weather Rev. 133, 1155–1174. https://doi.org/10.1175/MWR2906.1

Ray, P.A., Brown, C.M., 2015. Confronting Climate Uncertainty in Water Resources Planning and Project Design: The Decision Tree Framework. The World Bank. https://doi.org/10.1596/978-1-4648-0477-9

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – A framework and guidance. Environ. Model. Softw. 22, 1543–1556. https://doi.org/10.1016/J.ENVSOFT.2007.02.004

Reggiani, P., Coccia, G., Mukhopadhyay, B., 2016. Predictive Uncertainty Estimation on a Precipitation and Temperature Reanalysis Ensemble for Shigar Basin, Central Karakoram. Water 8, 263. https://doi.org/10.3390/w8060263

Reichert, P., 2012. Conceptual and Practical Aspects of Quantifying Uncertainty in Environmental Modelling and Decision Support, in: International Environmental Modelling and Software Society (IEMSs). Leipzig, Germany.

Reichert, P., Langhans, S.D., Lienert, J., Schuwirth, N., 2015. The conceptual foundation of environmental decision support. J. Environ. Manage. 154, 316–332. https://doi.org/10.1016/J.JENVMAN.2015.01.053

Reichert, P., Mieleitner, J., 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resour. Res. 45, W10402. https://doi.org/10.1029/2009WR007814

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resour. Res. 46, W05521. https://doi.org/10.1029/2009WR008328

Robert, C.P., 2016. Approximate Bayesian Computation: A Survey on Recent Results, in: Monte Carlo and Quasi-Monte Carlo Methods. Springer, Cham, pp. 185–205. https://doi.org/10.1007/978-3-319-33507-0_7

Rodriguez-Lloveras, X., Bussi, G., Francés, F., Rodriguez-Caballero, E., Solé-Benet, A., Calle, M., Benito, G., 2015. Patterns of runoff and sediment production in response to land-use changes in an ungauged Mediterranean catchment. J. Hydrol. 531, 1054–1066. https://doi.org/10.1016/J.JHYDROL.2015.11.014

Romero-Cuellar, J., Abbruzzo, A., Adelfio, G., Francés, F., 2019. Hydrological post-processing based on approximate Bayesian computation (ABC). Stoch Env. Res Risk Assess 33, 1361–1373. https://doi.org/10.1007/s00477-019-01694-y

Romero-Cuéllar, J., Buitrago-Vargas, A., Quintero-Ruiz, T., Francés, F., 2018. Simulación hidrológica de los impactos potenciales del cambio climático en la cuenca hidrográfica del río Aipe, en Huila, Colombia. Ribagua 1–16.

https://doi.org/10.1080/23863781.2018.1454574

Roundy, J.K., Duan, Q., Schaake, J., 2018. Hydrological Predictability, Scales, and Uncertainty Issues, in: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–29. https://doi.org/10.1007/978-3-642-40457-3_8-1

Ruiz-Pérez, G., Medici, C., Latron, J., Llorens, P., Gallart, F., Francés, F., 2016. Investigating the behaviour of a small Mediterranean catchment using three different hydrological models as hypotheses. Hydrol. Process. 30, 2050–2062. https://doi.org/10.1002/hyp.10738

Ruiz-Villanueva, V., Stoffel, M., Bussi, G., Francés, F., Bréthaut, C., 2015. Climate change impacts on discharges of the Rhone River in Lyon by the end of the twenty-first century: model results and implications. Reg. Environ. Chang. 15, 505–515. https://doi.org/10.1007/s10113-014-0707-8

Schefzik, R., Thorarinsdottir, T.L., Gneiting, T., 2013. Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling. Stat. Sci. 28, 616–640. https://doi.org/10.2307/43288438

Schenk, N.J., Lensink, S.M., 2007. Communicating uncertainty in the IPCC's greenhouse gas emissions scenarios. Clim. Change 82, 293–308. https://doi.org/10.1007/s10584-006-9194-3

Schepen, A., Wang, Q.J., 2015. Model averaging methods to merge operational statistical and dynamic seasonal streamflow forecasts in Australia. Water Resour. Res. 51, 1797–1812. https://doi.org/10.1002/2014WR016163

Schepen, A., Zhao, T., Wang, Q.J., Robertson, D.E., 2018. A Bayesian modelling method for post-processing daily sub-seasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 Australian catchments. Hydrol. Earth Syst. Sci. 22, 1615–1628. https://doi.org/10.5194/hess-22-1615-2018

Schoups, G., van de Giesen, N.C., Savenije, H.H.G., 2008. Model complexity control for hydrologic prediction. Water Resour. Res. 44, W00B03. https://doi.org/10.1029/2008WR006836

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resour. Res. 46, W10531. https://doi.org/10.1029/2009WR008933

Shafii, M., Tolson, B., Matott, L.S., 2014. Uncertainty-based multi-criteria calibration of rainfall-runoff models: a comparative study. Stoch. Environ.

Res. Risk Assess. 28, 1493–1510. https://doi.org/10.1007/s00477-014-0855-x

Sharma, S., Siddique, R., Reed, S., Ahnert, P., Mendoza, P., Mejia, A., 2018. Relative effects of statistical preprocessing and postprocessing on a regional hydrological ensemble prediction system. Hydrol. Earth Syst. Sci. 22, 1831–1849. https://doi.org/10.5194/hess-22-1831-2018

Siegert, S., Stephenson, D.B., 2019. Forecast Recalibration and Multimodel Combination, in: Sub-Seasonal to Seasonal Prediction. Elsevier, pp. 321–336. https://doi.org/10.1016/B978-0-12-811714-9.00015-2

Sikorska, A.E., Montanari, A., Koutsoyiannis, D., 2015. Estimating the Uncertainty of Hydrological Predictions through Data-Driven Resampling Techniques. J. Hydrol. Eng. 20, A4014009. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000926

Sisson, S.A., Fan, Y., Tanaka, M.M., 2007. Sequential Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. U. S. A. 104, 1760–5. https://doi.org/10.1073/pnas.0607208104

Sivapalan, M., 2018. From engineering hydrology to Earth system science: milestones in the transformation of hydrologic science. Hydrol. Earth Syst. Sci. 22, 1665–1693. https://doi.org/10.5194/hess-22-1665-2018

Smith, T., Marshall, L., Sharma, A., 2015. Modeling residual hydrologic errors with Bayesian inference. J. Hydrol. 528, 29–37.

Smith, T., Marshall, L., Sharma, A., 2014. Predicting hydrologic response through a hierarchical catchment knowledgebase: A Bayes empirical Bayes approach. Water Resour. Res. 50, 1189–1204. https://doi.org/10.1002/2013WR015079

Smith, T.J., Marshall, L.A., 2008. Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. Water Resour. Res. 44. https://doi.org/10.1029/2007WR006705

Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model uncertainty using machine learning techniques. Water Resour. Res. 45, W00B11. https://doi.org/10.1029/2008WR006839

Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. Water Resour. Res. 16, 430–442. https://doi.org/10.1029/WR016i002p00430

Stainforth, D.A., Downing, T.E., Washington, R., Lopez, A., New, M., 2007. Issues in the interpretation of climate model ensembles to inform decisions. Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci. 365.

Stakhiv, E., Stewart, B., 2010. Needs for Climate Information in Support of Decision-Making in the Water Sector. Procedia Environ. Sci. 1, 102–119. https://doi.org/10.1016/J.PROENV.2010.09.008

Steinschneider, S., Polebitski, A., Brown, C., Letcher, B.H., 2012. Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change. Water Resour. Res. 48, W11525. https://doi.org/10.1029/2011WR011318

Stoll, S., Hendricks Franssen, H.J., Butts, M., Kinzelbach, W., 2011. Analysis of the impact of climate change on groundwater related hydrological fluxes: a multi-model approach including different downscaling methods. Hydrol. Earth Syst. Sci. 15, 21–38. https://doi.org/10.5194/hess-15-21-2011

Sun, R., Yuan, H., Liu, X., 2017. Effect of heteroscedasticity treatment in residual error models on model calibration and prediction uncertainty estimation. J. Hydrol. 554, 680–692. https://doi.org/10.1016/J.JHYDROL.2017.09.041

Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P., 1997. Inferring Coalescence Times From DNA Sequence Data. Genetics 145, 505–518.

Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci. 365.

Thomas, H., 1981. Improved methods for national water assessment, water resources contract: WR15249270. Cambridge.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. Water Resour. Res. 45, W00B14. https://doi.org/10.1029/2008WR006825

Tian, Y., Nearing, G.S., Peters-Lidard, C.D., Harrison, K.W., Tang, L., Tian, Y., Nearing, G.S., Peters-Lidard, C.D., Harrison, K.W., Tang, L., 2016. Performance Metrics, Error Modeling, and Uncertainty Quantification. Mon. Weather Rev. 144, 607–613. https://doi.org/10.1175/MWR-D-15-0087.1

Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. Int. J. River Basin Manag. 6, 123–137. https://doi.org/10.1080/15715124.2008.9635342

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.H., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6, 187–202. https://doi.org/10.1098/RSIF.2008.0172

Tran, M.-N., Nott, D.J., Kohn, R., 2017. Variational Bayes With Intractable Likelihood. J. Comput. Graph. Stat. 26, 873–882. https://doi.org/10.1080/10618600.2017.1330205

Turner, B.M., Van Zandt, T., 2012. A tutorial on approximate Bayesian computation. J. Math. Psychol. 56, 69–85. https://doi.org/10.1016/j.jmp.2012.02.005

van Andel, S.J., Weerts, A., Schaake, J., Bogner, K., 2013. Post-processing hydrological ensemble predictions intercomparison experiment. Hydrol. Process. 27, 158–161. https://doi.org/10.1002/hyp.9595

van der Linden, P., Mitchell, J., 2009. ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project. FitzRoy Road.

Van Der Waerden, B.L., 1953. Order Tests for the Two-Sample Problem and their Power. Indag. Math. 56, 80. https://doi.org/10.1016/S1385-7258(53)50012-5

van Oijen, M., 2017. Bayesian Methods for Quantifying and Reducing Uncertainty and Error in Forest Models. Curr. For. Reports 3, 269–280. https://doi.org/10.1007/s40725-017-0069-9

Veldkamp, T.I.E., Wada, Y., Aerts, J.C.J.H., Döll, P., Gosling, S.N., Liu, J., Masaki, Y., Oki, T., Ostberg, S., Pokhrel, Y., Satoh, Y., Kim, H., Ward, P.J., 2017. Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century. Nat. Commun. 8, 15697. https://doi.org/10.1038/ncomms15697

Vélez, J.J., Puricelli, M., López Unzu, F., Francés, F., 2009. Parameter extrapolation to ungauged basins with a hydrological distributed model in a regional framework. Hydrol. Earth Syst. Sci. 13, 229–246. https://doi.org/10.5194/hess-13-229-2009

Vogel, R.M., 2017. Stochastic watershed models for hydrologic risk management. Water Secur. 1, 28–35. https://doi.org/10.1016/J.WASEC.2017.06.001

Vogel, R.M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P.K., Hooper, R.P., Matalas, N.C., 2015. Hydrology: The interdisciplinary science of water.

Water Resour. Res. 51, 4409–4430. https://doi.org/10.1002/2015WR017049

von Trentini, F., Leduc, M., Ludwig, R., 2019. Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble. Clim. Dyn. 1–17. https://doi.org/10.1007/s00382-019-04755-8

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 467, 555–561. https://doi.org/10.1038/nature09440

Vrugt, J.A., Massoud, E.C., 2018. Uncertainty Quantification of Complex System Models: Bayesian Analysis, in: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–74. https://doi.org/10.1007/978-3-642-40457-3_27-1

Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resour. Res. 43, W01411. https://doi.org/10.1029/2005WR004838

Vrugt, J.A., Sadegh, M., 2013. Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resour. Res. 49, 4335–4345. https://doi.org/10.1002/wrcr.20354

Wagener, T., Gupta, H. V., 2005. Model identification for hydrological forecasting under uncertainty. Stoch. Environ. Res. Risk Assess. 19, 378–387. https://doi.org/10.1007/s00477-005-0006-5

Wagener, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H. V., Kumar, P., Rao, P.S.C., Basu, N.B., Wilson, J.S., 2010. The future of hydrology: An evolving science for a changing world. Water Resour. Res. 46. https://doi.org/10.1029/2009WR008906

Wang, C., Duan, Q., Tong, C.H., Di, Z., Gong, W., 2016. A GUI platform for uncertainty quantification of complex dynamical models. Environ. Model. Softw. 76, 1–12. https://doi.org/10.1016/J.ENVSOFT.2015.11.004

Wang, Q.J., Robertson, D.E., Chiew, F.H.S., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. Water Resour. Res. 45. https://doi.org/10.1029/2008WR007355

Wang, Q.J., Shrestha, D.L., Robertson, D.E., Pokhrel, P., 2012. A log-sinh transformation for data normalization and variance stabilization. Water Resour. Res. 48. https://doi.org/10.1029/2011WR010973

Weaver, C.P., Lempert, R.J., Brown, C., Hall, J.A., Revell, D., Sarewitz, D., 2013. Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks. Wiley Interdiscip. Rev. Clim. Chang. 4, 39–60. https://doi.org/10.1002/wcc.202

Weerts, A.H., Winsemius, H.C., Verkade, J.S., 2011. Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). Hydrol. Earth Syst. Sci 15, 255–265. https://doi.org/10.5194/hess-15-255-2011

Wheater, H.S., Gober, P., 2015. Water security and the science agenda. Water Resour. Res. 51, 5406–5424. https://doi.org/10.1002/2015WR016892

Wilby, R.L., Harris, I., 2006. A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. Water Resour. Res. 42, W02419. https://doi.org/10.1029/2005WR004065

Wild, M., Liepert, B., 2010. The Earth radiation balance as driver of the global hydrological cycle. Environ. Res. Lett. 5, 025203. https://doi.org/10.1088/1748-9326/5/2/025203

Winsemius, H.C., Dutra, E., Engelbrecht, F.A., Archer Van Garderen, E., Wetterhall, F., Pappenberger, F., Werner, M.G.F., 2014. The potential value of seasonal forecasts in a changing climate in southern Africa. Hydrol. Earth Syst. Sci. 18, 1525–1538. https://doi.org/10.5194/hess-18-1525-2014

Woldemeskel, F., McInerney, D., Lerat, J., Thyer, M., Kavetski, D., Shin, D., Tuteja, N., Kuczera, G., 2018. Evaluating post-processingapproaches for monthly and seasonal streamflow forecasts. Hydrol. Earth Syst. Sci. 22, 6257–6278. https://doi.org/10.5194/hess-2018-214

Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Adv. Water Resour. 30, 1756–1774. https://doi.org/10.1016/J.ADVWATRES.2007.01.005

Ye, A., Duan, Q., Yuan, X., Wood, E.F., Schaake, J., 2014. Hydrologic post-processing of MOPEX streamflow simulations. J. Hydrol. 508, 147–156. https://doi.org/10.1016/j.jhydrol.2013.10.055

Yoon, S., Cho, W., Heo, J.-H., Kim, C.E., 2010. A full Bayesian approach to generalized maximum likelihood estimation of generalized extreme value distribution. Stoch. Environ. Res. Risk Assess. 24, 761–770. https://doi.org/10.1007/s00477-009-0362-7

Zadeh, L.A., 2005. Toward a generalized theory of uncertainty (GTU)—an outline. Inf. Sci. (Ny). 172, 1–40. https://doi.org/10.1016/J.INS.2005.01.017

Zhang, X., Zhao, K., 2012. Bayesian Neural Networks for Uncertainty Analysis of Hydrologic Modeling: A Comparison of Two Schemes. Water Resour. Manag. 26, 2365–2382. https://doi.org/10.1007/s11269-012-0021-5

Zhang, Z., Wagener, T., Reed, P., Bhushan, R., 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resour. Res. 44, W00B04. https://doi.org/10.1029/2008WR006833

Zhao, L., Duan, Q., Schaake, J., Ye, A., Xia, J., 2011. A hydrologic post-processor for ensemble streamflow predictions. Adv. Geosci. 29, 51–59. https://doi.org/10.5194/adgeo-29-51-2011

Zhao, T., Bennett, J.C., Wang, Q.J., Schepen, A., Wood, A.W., Robertson, D.E., Ramos, M.-H., Zhao, T., Bennett, J.C., Wang, Q.J., Schepen, A., Wood, A.W., Robertson, D.E., Ramos, M.-H., 2017. How Suitable is Quantile Mapping For Postprocessing GCM Precipitation Forecasts? J. Clim. 30, 3185–3196. https://doi.org/10.1175/JCLI-D-16-0652.1

Zhu, W., Marin, J.M., Leisen, F., 2016. A Bootstrap Likelihood Approach to Bayesian Computation. Aust. N. Z. J. Stat. 58, 227–244. https://doi.org/10.1111/anzs.12156