Article

Quantile-Based Hydrological Modelling

Hristos Tyralis [1,2,*] and Georgia Papacharalampous [3]

1   Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 157 80 Zografou, Greece

2   Air Force Projects Authority, Hellenic Air Force, Mesogion Avenue 227–231, 15 561 Cholargos, Greece

3   Department of Engineering, Roma Tre University, Via V. Volterra 62, 00146 Rome, Italy; papacharalampous.georgia@gmail.com

*   Correspondence: montchrister@gmail.com

Abstract: Predictive uncertainty in hydrological modelling is quantified by using post-processing or Bayesian-based methods. The former methods are not straightforward and the latter ones are not distribution-free (i.e., assumptions on the probability distribution of the hydrological model's output are necessary). To alleviate possible limitations related to these specific attributes, in this work we propose the calibration of the hydrological model by using the quantile loss function. By following this methodological approach, one can directly simulate pre-specified quantiles of the predictive distribution of streamflow. As a proof of concept, we apply our method in the frameworks of three hydrological models to 511 river basins in the contiguous US. We illustrate the predictive quantiles and show how an honest assessment of the predictive performance of the hydrological models can be made by using proper scoring rules. We believe that our method can help towards advancing the field of hydrological uncertainty.

Keywords: big data; catchment models; hydrological models; hydrological uncertainty; large-sample hydrology; point predictions; predictive uncertainty; quantile loss; quantile regression; uncertainty

1. Introduction

One purpose of hydrological models is to provide predictions of streamflow [1]. A split-sample scheme is implemented to assess the quality of predictions. The hydrological model is calibrated in the first segment of the sample, while the second segment is used for validation (testing) [2,3]. The validation procedure should be applied by using information not available during the calibration procedure, while the model's performance can be assessed using some quantitative criteria (see [3] for guidelines on hydrological model validation). A key component of a hydrological model is its objective function, also called loss function, cost function or scoring function. The choice of objective function depends on the purpose of the modelling, which in most cases is to provide predictions close to the observations [1]. To this end, objective functions, such as the squared error or the Nash–Sutcliffe efficiency [4], are implemented.

Besides predicting the observations as closely (i.e., accurately) as possible, quantifying predictive uncertainty is also of high interest [5,6]. Predictive uncertainty is defined here as the uncertainty of hydrological predictions (simulations) [7] and can be quantified by a probability function. Quantification of the uncertainty of hydrological predictions is closely related to the 20th unsolved problem in hydrology: "*How can we disentangle and reduce model structural/parameter/input uncertainty in hydrological prediction?*" [8].

Numerous methods have been proposed and implemented for quantifying predictive uncertainty by providing predictive quantiles or full predictive distribution, including:

a.   Post-processing (two-stage) techniques using stochastic or regression-based methods (including machine learning ones). These techniques model residual errors. Several relevant examples can be found in the literature; see [9–37] and the review by [38],

including quantile regression, which is based on the quantile loss function. Further classification of the methods is possible, depending on specific attributes, but this is out of scope here.

b.   Monte Carlo schemes; see [39–41].
c.   Joint inference methods that are mostly based on Bayesian techniques; see [16,42,43].
d.   Ensembles of predictions (see [44]) that may need to be post-processed as well [9,45–47].
e.   Combinations of the above schemes [41,48,49].

Post-processing techniques are popular in forecasting applications, in which a deterministic (point) forecast by the hydrological model is delivered and the modeller conditions the predictand on the forecast as well as past input and output observations to obtain a predictive distribution [19]. In simulation mode, post-processing techniques are also applicable. Joint inference techniques are popular in hydrological simulation and skip the two-stage procedure, e.g., using Bayesian methods [16]. In both cases (simulation and forecasting), the predictive distribution of the output [50] (p. 22) is delivered and can be used to quantify predictive uncertainty.

Of special interest in hydrological modelling is the prediction of specific low or high quantiles of streamflow. This problem can take the form of one in which the modeller receives a directive in the form of a specific statistical functional, i.e., a quantile of the predictive distribution [51]. In this case, it is critical that the objective function "*is consistent for it, in the sense that the expected score is minimized when following the directive*" ([51], see the definition of consistency of scoring functions in Methods Section 2).

Following the above conversation, the aim of this work is to solve a problem that we define as follows: "*A hydrological modeller receives a directive to predict a quantile of the distribution of streamflow. How could she/he adapt its model to do so?*". To this end, we propose the use of quantile loss [52] as the objective function of the hydrological model at the requested (by the end user) quantile levels.

We note here that compared to the previous mainstream approaches to quantifying predictive uncertainty we differentiate in the following (while implications are discussed in detail in the Discussion and Conclusions in Sections 6 and 7, respectively):

a.   Post-processing approaches: The quantile loss function is directly implemented by the hydrological model. On the other hand, post-processing methods model the residuals of the fitted hydrological model (that have been obtained by implementing a squared error type loss function) at a second stage (i.e., after applying the hydrological models), using a statistical or machine learning method.

b.   Monte Carlo of Bayesian joint inference approaches: These methods are applied directly to the hydrological model; therefore, there is some resemblance with our approach. In this case, the differences between the two approaches (Bayesian joint inference and our proposed approach) are identical with those identified in the statistical literature regarding the differences between Bayesian statistics and quantile regression modelling. These differences are thoroughly discussed in the Discussion in Section 6.

Statistical modelling based on the quantile loss function is frequently covered in the statistical literature and is well received by practitioners, e.g., in the optimisation of linear-in-parameters models (e.g., see the book by [53]), neural networks [54], random forests (e.g., see the review by [55]) and boosting algorithms (e.g., see the review by [56]). Therefore, we believe that it will also be of practical interest to hydrologists. Calibration of a hydrological model with the quantile loss function has also been proposed by [57,58] in the context of model selection and model structure deficiency assessment; however, the value of the quantile loss function for predictive uncertainty assessment has been minimally examined, while here we frame assessments of predictive uncertainty in a formal framework.

The remainder of the paper is structured as follows. The methods, with emphasis on concepts related to the quantile loss function, are presented in Section 2. In the same section, the suite of the three lumped GR (Génie Rural) hydrological models [59] used in the study is briefly presented. The data used to apply the hydrological models are presented

in Section 3. In brief, this data originates from 511 river basins in the CONUS (contiguous US). The dataset description is followed by a summary and implementation directives for the proposed framework, which are provided in Section 4. The results of the application of the hydrological models are presented in Section 5. Lastly, the paper closes with the Discussion and Conclusions in Sections 6 and 7, respectively.

2. Methods

Here, we present the concept of the quantile loss function whose use is proposed in the manuscript for supporting the application of a hydrological model to directly predict the predictive quantiles of interest. An outline of the properties of this loss function is also provided, followed by a brief presentation of the three lumped GR hydrological models.

2.1. Quantile Loss Function

A real-valued random variable $\underline{x}$ might be characterised by its distribution function F, defined by:

$$F(x):= P(\underline{x} \leq x), \tag{1}$$

Then, $F^{-1}(a)$ is defined by:

$$F^{-1}(a):= \inf\{x\colon F(x) \geq a\}, \tag{2}$$

$F^{-1}(a)$ is referred to as the ath quantile of x, while inf denotes the infimum of a set of real numbers. For instance, $F^{-1}(1/2)$ is the median or 0.50th quantile. In regression modelling, one minimises the sum of absolute errors to estimate the median of the conditional distribution. The natural question then is "*are there analogs for regression of the other quantiles?*" [53] (p. 5). The idea, elaborated by [52], is to apply the quantile loss function defined by Equation (3), instead of using the absolute error function:

$$L(r; x, a):= (r - x)\,(\mathbf{1}(x \leq r) - a), \tag{3}$$

Here, $\mathbf{1}(\cdot)$ denotes the indicator function, x is the materialisation of the variable $\underline{x}$, a is the quantile level of interest and r is the respective predictive quantile. For $a = 1/2$, Equation (3) reduces to:

$$L(r; x, 1/2) = |r - x|/2, \tag{4}$$

which is half the absolute error function. Obviously, when a sample is provided, the objective is to minimise the average score, i.e., the quantile loss function averaged over a fixed set of observations, equal to $\sum_i L(r_i; x_i, a)/n$, $i \in \{1, \ldots, n\}$.

The quantile loss function, defined by Equation (3), is positive and negatively oriented, i.e., the objective is to minimise it. Figure 1 illustrates the quantile loss function for $x = 0$ and varying predictive quantiles rs at quantile levels 0.05 and 0.95 (to understand how this loss function works). The quantile loss function is asymmetric. At the 0.05 quantile level (and, in general, when $a < 1/2$), it assigns a penalty to r lower than the materialisation, compared to its symmetric predictive quantile. The quantile loss function equals to 0, when $r = x$.
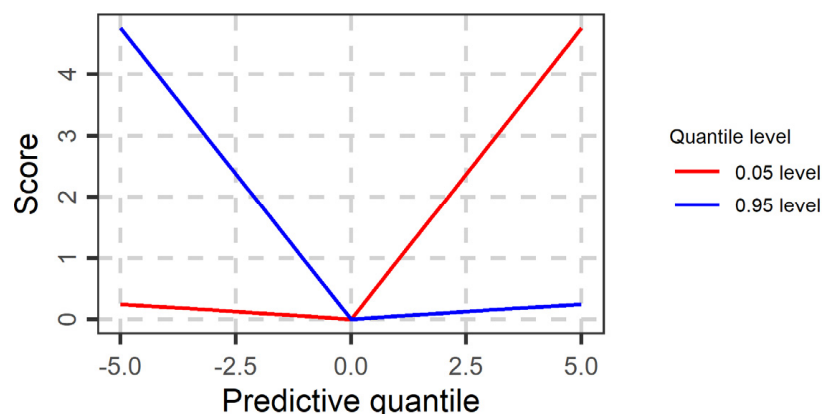
Figure 1. Illustration of the quantile score (value of the quantile loss function) at the quantile levels $a \in \{0.05, 0.95\}$ when $x = 0$ materialises and for varying predictive quantiles r (see Equation (3)).

2.2. Theoretical Properties of the Quantile Loss Function

We note that linear-in-parameters quantile regression (or simply quantile regression when referring to non-linear models) is based on the minimisation of the quantile loss function; therefore, in what follows, we refer to both the properties of quantile regression and quantile loss. A history of concepts related to quantile regression can be found in [60].

Following [51] and the intuitive explanation by [61], if a modeller "*is asked to report a certain functional of the predictive distribution, then a key requirement on the loss function is to be strictly consistent, in the sense that the expected loss or score is uniquely minimized if the directive asked for is followed*" [51]. It is proved by [51] that the quantile loss function is consistent for the quantiles of the predictive distribution.

In the context of probabilistic predictions, scoring rules are used to quantify predictive performance [62]. "*A scoring rule is proper if truth telling is an optimal strategy in expectation*" [63]. Scoring rules support the modeller in being honest about the assessment of his predictions [62]. It has been proved that the quantile loss function is a proper scoring rule [62].

Moving from the concepts of consistency and propriety, the quantile loss function has been the focus of intensive research, including optimisation algorithms for parameter estimation in quantile regression settings [64,65], goodness-of-fit processes [66], decomposition in reliability, resolution and uncertainty [67], and more.

2.3. Hydrological Models

The Génie Rural GR4J, GR5J and GR6J daily lumped hydrological models were used in the study by [59]. These models are widely used in hydrology, while their detailed description is out of the scope of the study. Their modular implementation in R programming language allows the minimisation of a customised loss function.

The GR hydrological models have evolved over time from the daily GR3J model with three parameters [68] to the GR4J model with four parameters [69] and the GR5J and GR6J models with five and six parameters, respectively [70]. Other versions of the GR models are available at monthly [71] and annual [72] time scales. The early history of the GR models can be found in [59].

The estimation of the parameters of the hydrological model is done using Michel's [73] optimisation algorithm, while the models here are calibrated and evaluated using the quantile loss function. The GR models have some interesting properties that differentiate them from one another. The additional parameter of the GR5J model, compared to the GR4J model, considers more complex inter-catchment water exchanges, while the GR6J model offers improvements in the simulation of low flows [59]. Presentations of the parameters of the models can be found in [69,70].

3. Data

We applied our method to daily hydrometeorological data from 511 small- to medium-sized river basins. These river basins represent most climate types over the CONUS, and their changes due to human influences are minimal. The data and detailed information about the river basins are available in the CAMELS dataset, which is fully described in [74–79]; see Figure 2 for the geographical locations of the river basins. Daily time series of minimum and maximum temperatures and precipitation are available for each river basin [79]. The same applies for daily runoff time series. Here, we focus on the 34-year period 1980–2013. The same 34-year period and river basins have been previously examined in the hydrological post-processing study by [33], as for them the available time series records are complete. Mean daily temperatures were computed for each river basin by averaging its minimum and maximum daily temperatures; this was done separately for each day included in the 34-year period. Lastly, a time series of daily potential evapotranspiration (PET) was estimated for each river basin by applying Oudin's formula [80] to the previously obtained daily mean temperature time series.
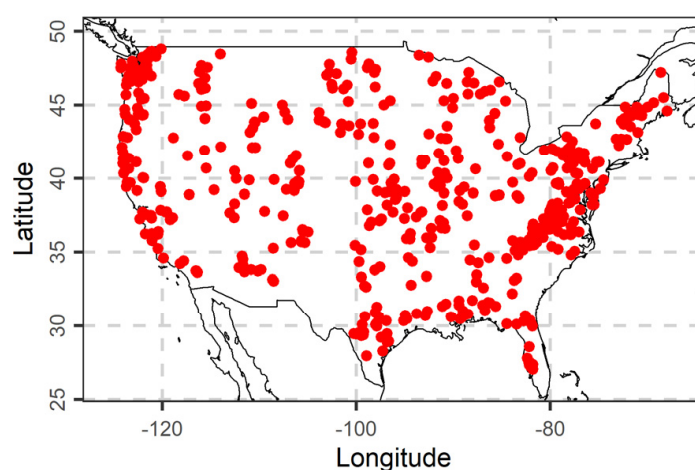

Figure 2. The 511 river basin stations over CONUS examined in the study.

4. Implementation and Key Components

Here, we present the key components and steps of the conducted large-scale application, thereby summarizing Sections 2 and 3. We applied three hydrological models, specifically the GR4J, GR5J and GR6J ones. The inputs to these models are daily precipitation and potential evapotranspiration time series, while their output is daily runoff time series. The period of interest is 1980–2013 and the dataset spans across 511 river basins in the CONUS. The `airGR` R package was used to implement the hydrological models after appropriate adaptations. The latter facilitate hydrological model calibration based on the quantile loss function.

For an arbitrary river basin, the adopted procedures were the following:

a. Define the 2-year period 1980–1981 as the warm-up period of the hydrological models.
b. Calibrate the hydrological models in the 16-year period 1982–1997 using: (a) the quantile loss function $L(r; x, a)$ at quantile levels $a \in \{0.025, 0.050, 0.100, 0.500, 0.900, 0.950, 0.975\}$; and (b) the squared error function. That equates to 8 (i.e., 7 + 1; number of loss functions) × 3 (number of hydrological models) = 24 sets of parameters at each river basin. These sets of parameters correspond to different model setups.
c. Simulate streamflow in the 16-year period 1998–2013 for each of the 24 model setups.

In what follows, the period 1998–2013 will be used for model testing purposes, while all results refer to the validation period. All methods were implemented in the R programming language using the contributed packages mentioned in Appendix A.

5. Results

To facilitate the desired understanding of the outputs of the quantile-based hydrological models, in Figure 3 we present the predicted streamflow at an arbitrary river basin and for an arbitrary two-year period. We note that the models simulated streamflow in the entire 16-year period 1998–2013, and restricting the presentation of the model outputs within an arbitrary two-year period is here made for illustration purposes only. Let us now focus, for example, on the output of the GR4J model and specifically on the case that this model is calibrated at the quantile level 0.500. The simulated streamflow is close to the observed one, although the match is not perfect, since results at specific cases are subject to randomness; therefore, large-sample experiments will be presented in the following for accurate assessments. To fully perceive this outcome, it is also important to recall that the use of the $L(r; x, 0.500)$ loss function is equivalent to using the absolute error function. Instead, when the GR4J model was calibrated with the $L(r; x, 0.025)$ loss function, it simulated the streamflow quantile at level 0.025, which is lower than the observed streamflow. Similar observations and interpretations can be provided for the remaining quantile levels and hydrological models. It is also relevant to note at this point that the 0.025 and 0.975 quantiles contain the observed streamflow.
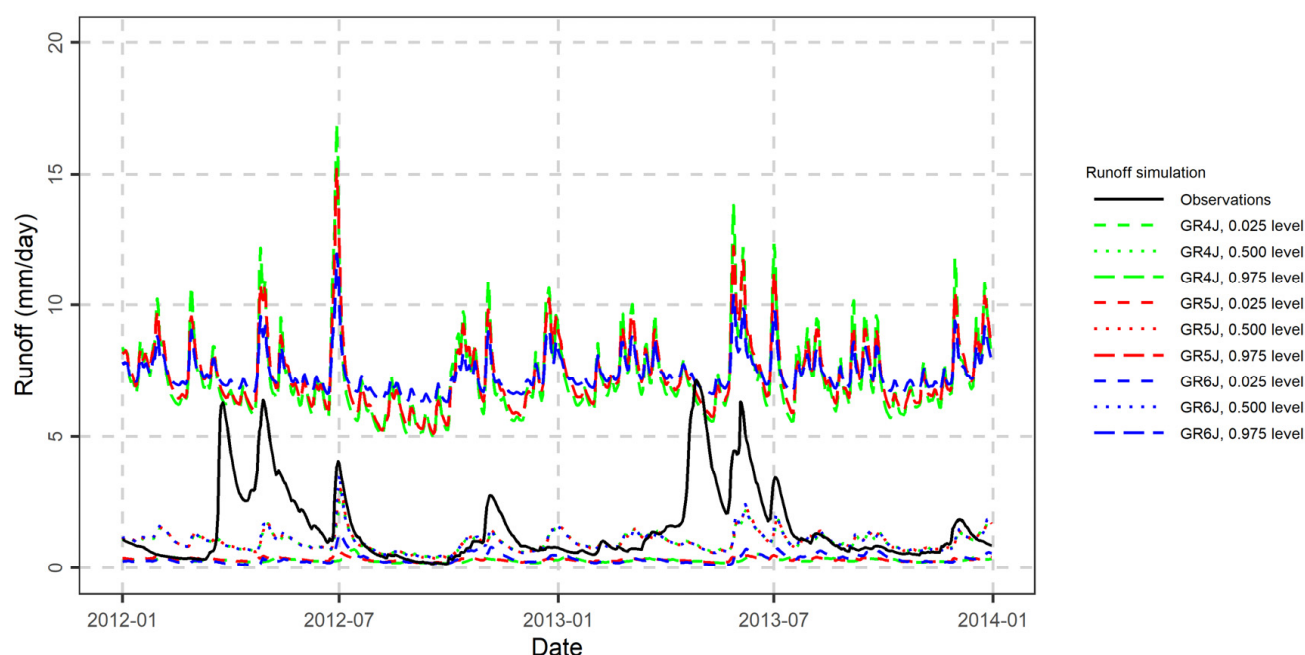

Figure 3. Illustration of observed streamflows and quantiles predicted by the GR4J, GR5J and GR6J hydrological models at the levels $a \in \{0.025, 0.500, 0.975\}$ for a two-year period at an arbitrary river basin.

The most frequently used objective functions in hydrological modelling are those of the squared-error type. To perceive the differences between the squared error and the $L(r; x, 1/2)$ loss functions, in Figure 4 we present the average score $L(r; x, 1/2)$ in the test period at each river basin for the predictions issued by the GR4J model and specifically for the case that this model is calibrated by using the $L(r; x, 1/2)$ loss function (x-axis) in comparison to the squared error function (y-axis). It is evident that when the calibration of the hydrological model is based on the $L(r; x, 1/2)$ loss function, one receives lower average scores in the test period compared to the case in which the calibration is based on the squared error loss function; recall also that the quantile loss function is negatively oriented; therefore, the lower the score, the better the prediction. On the other hand, it is evident that the average scores do not differ dramatically. This is reasonable given that both loss functions aim at simulating closely the observed streamflow.
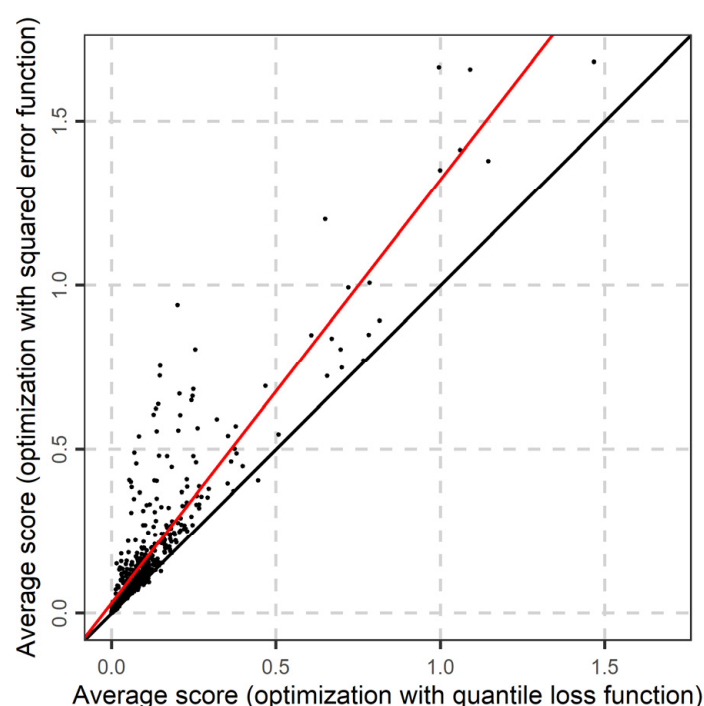
Figure 4. Scatterplot of the mean quantile scores at the level *a* = 0.500, as computed for the testing periods and separately for each of the 511 river basins, in the case that the quantile loss function at the level *a* = 0.500 is minimised (*x*-axis), and in the case that the squared error function is minimised (*y*-axis) for the calibration of the GR4J model.

When one performs a large-scale experiment to compare several models across a large number of river basins, illustrations such as those of Figure 3 cannot adequately support the assessment. Results should be summarised by using scaled scores. Unfortunately, quantile scores are scale-dependent; therefore, to facilitate a proper comparison a relative measure should be used. In such situations, a frequently used solution is to set a benchmark model, usually the simpler one in the experiments (which here is the GR4J model), and compare the performance of the remaining models with that of the benchmark. In particular, a relative $SCORE_{rel}$ measure can be defined by Equation (5) [81]; note the signs of the benchmark's $SCORE_{bench}$ and model's $SCORE_{model}$, reflecting that quantile loss functions are negatively oriented; therefore, improvements are obtained if $SCORE_{rel} > 0$:

$$SCORE_{rel} := (SCORE_{bench} - SCORE_{model})/SCORE_{bench}, \tag{5}$$

In Figure 5, we present the relative scores for the GR5J and GR6J models, against the average quantile score of the GR4J model. The average score refers to the loss function averaged over the test period, while it is computed separately for each set {hydrological model, calibration strategy (or targeted quantile level) and river basin}. In particular, the histogram of Figure 5a presents a sample of 7 (number of quantile levels) × 511 (number of river basins) = 3577 points, which are the relative scores of the GR5J model against the GR4J model at all the examined river basins and for the 7 examined quantile levels. The median improvement of the GR5J model against the GR4J model is 0.06%. On the other hand, the GR6J model is 1.55% worse compared to the GR4J model (see Figure 5b).
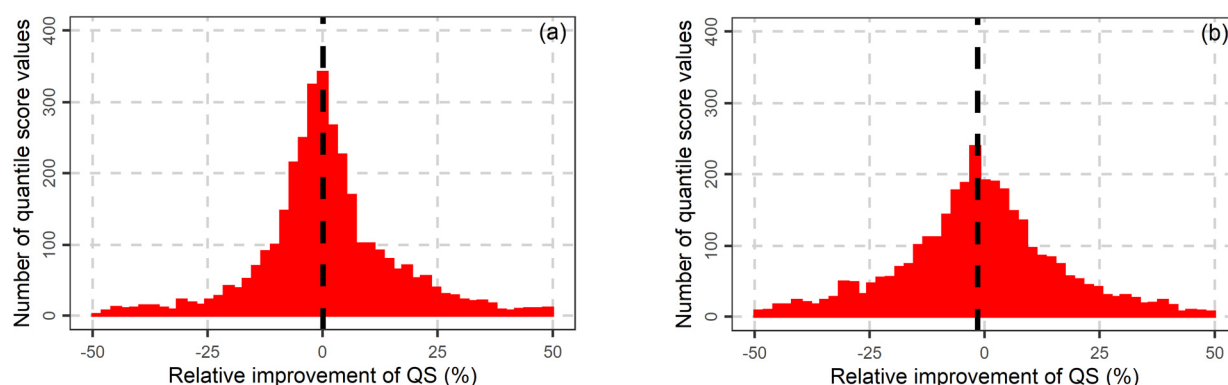
Figure 5. Histograms of the relative improvements (with red colour) and the median values of these relative improvements (with black dashed lines) against the performance of the GR4J hydrological model in terms of quantile score, as computed for all the 511 river basins and the 0.025, 0.050, 0.100, 0.500, 0.900, 0.950, 0.975 quantile levels, for the (**a**) GR5J and (**b**) GR6J hydrological models. Truncation at −50% and 50% has been applied for illustration purposes.

It is interesting to know if and how much each model improves over the performance of the benchmark at each quantile level. Figure 6 presents the median relative scores averaged over all river basins for the GR5J and GR6J models against the GR4J model. The performance of each model depends on the quantile level. For instance, the GR5J model improves over the performance of the GR4J model for quantile levels lower than 0.500, but its performance deteriorates for higher quantile levels.
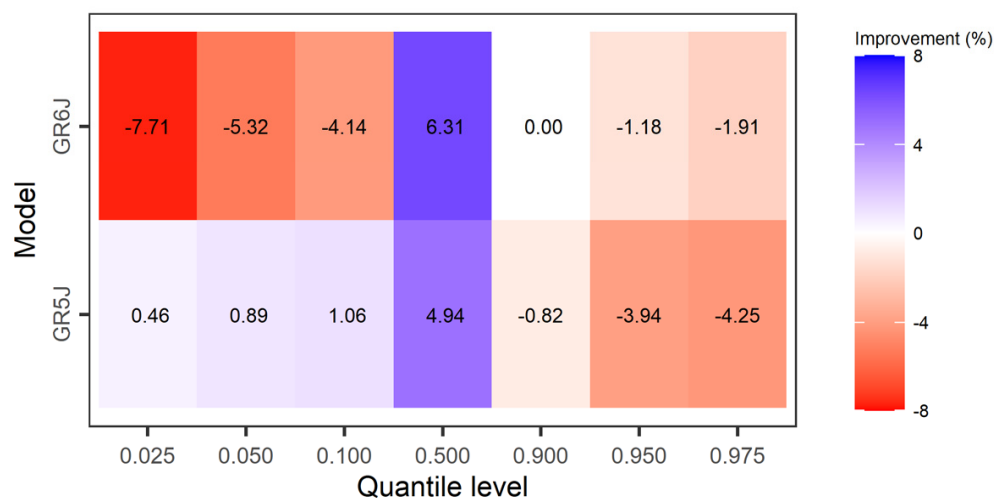

Figure 6. Heatmap of the median of the relative improvements summarizing the results for the 511 river basins of the performance of the GR5J and GR6J models against the performance of the GR4J hydrological model in terms of quantile scores.

Quantile scores are scale-dependent; therefore, intuition about the results across river basins, as presented above (i.e., in terms of average scores), is somewhat limited. In contrast, the coverage of the predictions issued by each model at each quantile level is a scale-independent measure. This measure counts the percentage of observations that are lower compared to the simulations at a specified quantile level. For instance, when simulating streamflow at quantile level 0.025, 2.50% of the observations will be lower than the simulated ones for a perfect model.

The median coverages of the predictions issued at all the examined river basins by the three models are provided in Figure 7. The simulations at high quantile levels seem to be better compared to those at lower quantile levels. This could be attributed to the fact that hydrological models are less skilled at simulating low and intermittent flows. Although coverages are intuitive, we note here that they are not consistent for reporting a certain

quantile of the predictive distribution of streamflow. Note that it is useful at this point to recall the definition of consistency (see Section 2.2). Therefore, they cannot support the modeller in being honest in the assessment of the hydrological models, in the sense which is explained in Section 2.2, and one should use proper scoring rules (e.g., the quantile score) to provide honest assessments. As an illustrative example, consider the case of obtaining predictive intervals (e.g., by combining predictive quantiles at 0.025 and 0.975 levels). One needs to have exact coverage, but also to obtain small-width intervals [82]; therefore, one has to rely on proper scoring rules.
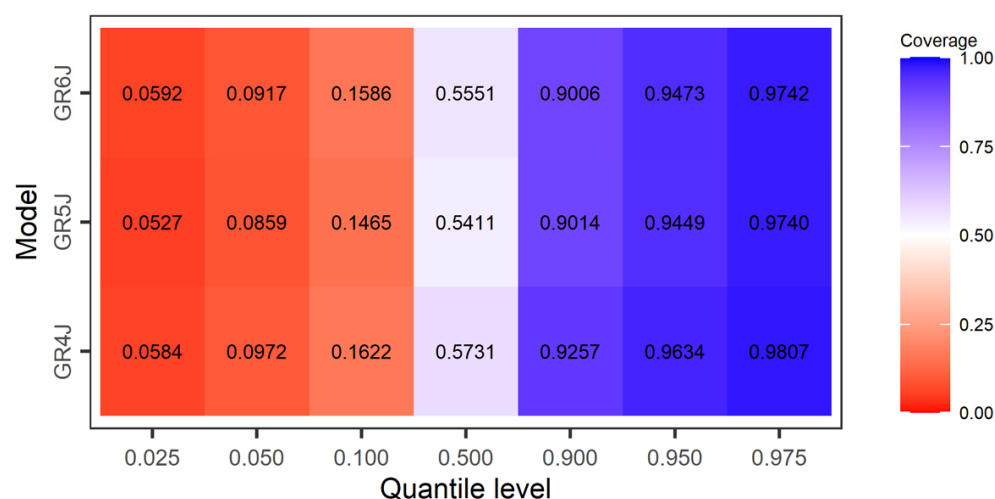

Figure 7. Heatmap of the median coverages of the predictions issued by the GR4J, GR5J and GR6J hydrological models at varying quantile levels summarizing the results for the 511 river basins.

6. Discussion

From a conceptual point of view, the differences between Bayesian approaches and our methodological approach to quantifying predictive uncertainty in hydrological modelling can be summarised by knowledge available in the field of statistics, as detailed in [83]. In particular, quantile regression (and therefore our methodological approach as well) is suitable in the following situations:

a. When one is interested in events at the "limits of probability".
b. When the conditional distribution does not follow a known distribution.
c. The possible presence of many outliers of the dependent variable (recall also that median regression is more robust compared to mean regression in the presence of outliers).
d. The presence of heteroscedasticity.

Obviously, streamflow attributes can be placed in the above situations; therefore, quantile-based hydrological modelling is an appropriate choice. Furthermore, compared to Bayesian approaches, quantile-based hydrological modelling is faster, as it practically constitutes an optimisation problem.

On the other hand, some drawbacks of our methodological approach are the following:

a. Estimating the parameters of the model is harder compared to Gaussian regression.
b. Inference on the parameters (e.g., the computation of confidence intervals) is complicated.
c. The possible presence of quantile crossing, i.e., estimated quantiles at higher levels might be lower than respective quantiles at lower levels.
d. The full conditional distribution is not available; although, the computation of multiple quantiles can substitute predictive distributions [61]. In this case, a drawback of the method is that it requires the estimation of a high number of sets of model parameters (one set for each quantile).

Compared to adopting post-processing approaches, directly changing the objective function is more straightforward; therefore, the predictions issued by the proposed method could be more interpretable. In particular, post-processing approaches resemble boosting algorithms with the difference that the base models of the former (i.e., the hydrological model and the post-processors) might be strong (i.e., they might provide good predictions), albeit they do not implement the loss function of interest. Given that boosting algorithms are designed to boost performance, it might be possible that post-processing will yield better or equivalent performance in practice. This could be investigated in the future by conducting independent large experiments (e.g., see the discussions in [84–87] on the value of big data experiments).

We also note that hydrological modellers are usually interested in low or high quantiles of streamflow. Although the model's structure can be customised to target such quantities, it is frequent to combine the model with a loss function suitable for modelling means (e.g., the squared error function or the Nash-Sutcliffe efficiency [88,89]). In such cases, one should not expect to obtain reliable estimates of quantiles of the predictive distribution, but rather customised estimates of the mean flow at low or high flow conditions.

Reporting some point predictions according to some vague request is a common practice that should be avoided; furthermore, it is not meaningful to evaluate the predictions using a set of scoring functions [51]. One should disclose the scoring function to the modeller or request a specific functional of the predictive distribution [51]. For instance, if the modeller receives a directive to report the mean functional, she/he should implement the squared error loss function which is consistent for this functional. Obviously, results could be reported for alternative scores, but they will be irrelevant for the task. Therefore, the comparison between the absolute error function and the squared error loss functions reported in Figure 4 is done for illustration purposes only and is not intended to prove that one function is better than the other.

Regarding the relative performance of the hydrological models at different quantile levels, more complex models provided higher performances for the 0.50 quantile; although, at other quantile levels a deterioration is observed with the exception of the GR5J model for quantile levels lower than 0.50. The reason for the performances of the models is unclear and could be investigated in a following study. However, it should be noted that GR hydrological models were designed with a focus on the mean functional, with more complex models delivering better predictions; however, these properties do not necessarily transfer to other quantile levels.

The quantile loss function can substitute the squared error function (or its equivalents) in hydrological models tailored to deliver forecasts; therefore, one can directly obtain quantile forecasts using a single model or post-process quantile simulation in the data assimilation procedure. Finally, similar methodological themes to those proposed in this work, including several ones for issuing point [90] and probabilistic [91] predictions of hydrological signatures, could be provided by exclusively using hydrological models instead of relying on data driven-ones. Other similar themes for improved quantile-based predictions are those combining multiple hydrological models and more [30].

7. Conclusions

In this paper, we proposed a new methodological approach to quantifying predictive uncertainty in hydrological modelling. The key idea is to calibrate the hydrological model by using the quantile loss function. By doing so, one can directly simulate the targeted quantiles of the predictive distribution of streamflow. Compared to post-processing techniques, our approach is straightforward and, compared to Bayesian techniques, it is distribution-free and faster.

To demonstrate the effectiveness of the new method and its wide applicability far from computational limitations, we applied it to 511 river basins in the CONUS in the wider frameworks of three hydrological models. With this large-scale application, we showed how the performances of different hydrological models can be compared with respect

to simulating predictive quantiles at pre-defined levels. Considering its simplicity, large applicability and other advantages, we believe that our method opens new avenues in the field of hydrological modelling, with its possible extensions including those quantifying uncertainty when predicting streamflow at ungauged basins.

Author Contributions: H.T. and G.P. contributed equally to the work. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CAMELS dataset, is fully described in [74–79].

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

Appendix A

The computations and visualisations were conducted in `R` Programming Language [92] by using the following packages: `airGR` [59,93], `data.table` [94], `devtools` [95], `gdata` [96], `knitr` [97–99], `rmarkdown` [100], `stringi` [101] and `tidyverse` [102,103].

References

1. Solomatine, D.P.; Wagener, T. 2.16-Hydrological Modeling. In *Treatise on Water Science*; Wilderer, P., Ed.; Elsevier: Amsterdam, The Netherlands, 2011; pp. 435–457. [CrossRef]
2. Klemeš, V. Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **1986**, *31*, 13–24. [CrossRef]
3. Biondi, D.; Freni, G.; Iacobellis, V.; Mascaro, G.; Montanari, A. Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Phys. Chem. Earth Parts A/B/C* **2012**, *42*, 70–76. [CrossRef]
4. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
5. Todini, E. Hydrological catchment modelling: Past, present and future. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 468–482. [CrossRef]
6. Ju, J.; Dai, H.; Wu, C.; Hu, B.X.; Ye, M.; Chen, X.; Gui, D.; Liu, H.; Zhang, J. Quantifying the uncertainty of the future hydrological impacts of climate change: Comparative analysis of an advanced hierarchical sensitivity in humid and semiarid basins. *J. Hydrometeorol.* **2021**, *22*, 601–621. [CrossRef]
7. Montanari, A. 2.17-Uncertainty of Hydrological Predictions. In *Treatise on Water Science*; Wilderer, P., Ed.; Elsevier: Amsterdam, The Netherlands, 2011; pp. 459–478. [CrossRef]
8. Blöschl, G.; Bierkens, M.F.P.; Chambel, A.; Cudennec, C.; Destouni, G.; Fiori, A.; Kirchner, J.W.; McDonnell, J.J.; Savenije, H.H.G.; Sivapalan, M.; et al. Twenty-three Unsolved Problems in Hydrology (UPH)–A community perspective. *Hydrol. Sci. J.* **2019**, *64*, 1141–1158. [CrossRef]
9. Biondi, D.; Todini, E. Comparing hydrological postprocessors including ensemble predictions into full predictive probability distribution of streamflow. *Water Resour. Res.* **2018**, *54*, 9860–9882. [CrossRef]
10. Bock, A.R.; Farmer, W.H.; Hay, L.E. Quantifying uncertainty in simulated streamflow and runoff from a continental-scale monthly water balance model. *Adv. Water Resour.* **2018**, *122*, 166–175. [CrossRef]
11. Bogner, K.; Pappenberger, F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resour. Res.* **2011**, *47*, W07524. [CrossRef]
12. Bogner, K.; Pappenberger, F.; Cloke, H.L. Technical note: The normal quantile transformation and its application in a flood forecasting system. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 1085–1094. [CrossRef]
13. Bogner, K.; Liechti, K.; Zappa, M. Post-processing of stream flows in Switzerland with an emphasis on low flows and floods. *Water* **2016**, *8*, 115. [CrossRef]
14. Bogner, K.; Liechti, K.; Zappa, M. Technical note: Combining quantile forecasts and predictive distributions of streamflows. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 5493–5502. [CrossRef]
15. Dogulu, N.; López López, P.; Solomatine, D.P.; Weerts, A.H.; Shrestha, D.L. Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 3181–3201. [CrossRef]

16. Evin, G.; Thyer, M.; Kavetski, D.; McInerney, D.; Kuczera, G. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resour. Res.* **2014**, *50*, 2350–2375. [CrossRef]
17. Krzysztofowicz, R. Markovian forecast processes. *J. Am. Stat. Assoc.* **1987**, *82*, 31–37. [CrossRef]
18. Krzysztofowicz, R. Transformation and normalization of variates with specified distributions. *J. Hydrol.* **1997**, *1997*, 286–292. [CrossRef]
19. Krzysztofowicz, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **1999**, *35*, 2739–2750. [CrossRef]
20. Krzysztofowicz, R. The case for probabilistic forecasting in hydrology. *J. Hydrol.* **2001**, *249*, 2–9. [CrossRef]
21. Krzysztofowicz, R. Bayesian system for probabilistic river stage forecasting. *J. Hydrol.* **2002**, *268*, 16–40. [CrossRef]
22. Krzysztofowicz, R.; Kelly, K.S. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* **2000**, *36*, 3265–3277. [CrossRef]
23. Li, D.; Marshall, L.; Liang, Z.; Sharma, A.; Zhou, Y. Characterizing distributed hydrological model residual errors using a probabilistic long short-term memory network. *J. Hydrol.* **2021**, *603*, 126888. [CrossRef]
24. Li, D.; Marshall, L.; Liang, Z.; Sharma, A.; Zhou, Y. Bayesian LSTM with stochastic variational inference for estimating model uncertainty in process-based hydrological models. *Water Resour. Res.* **2021**, *57*, e2021WR029772. [CrossRef]
25. López López, P.; Verkade, J.S.; Weerts, A.H.; Solomatine, D.P. Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn River: A comparison. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 3411–3428. [CrossRef]
26. Montanari, A.; Brath, A. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.* **2004**, *40*, W01106. [CrossRef]
27. Montanari, A.; Grossi, G. Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resour. Res.* **2008**, *44*, W00B08. [CrossRef]
28. Papacharalampous, G.; Tyralis, H.; Langousis, A.; Jayawardena, A.W.; Sivakumar, B.; Mamassis, N.; Montanari, A.; Koutsoyiannis, D. Probabilistic hydrological post-processing at scale: Why and how to apply machine-learning quantile regression algorithms. *Water* **2019**, *11*, 2126. [CrossRef]
29. Papacharalampous, G.; Koutsoyiannis, D.; Montanari, A. Quantification of predictive uncertainty in hydrological modelling by harnessing the wisdom of the crowd: Methodology development and investigation using toy models. *Adv. Water Resour.* **2020**, *136*, 103471. [CrossRef]
30. Papacharalampous, G.; Tyralis, H.; Koutsoyiannis, D.; Montanari, A. Quantification of predictive uncertainty in hydrological modelling by harnessing the wisdom of the crowd: A large-sample experiment at monthly timescale. *Adv. Water Resour.* **2020**, *136*, 103470. [CrossRef]
31. Seo, D.J.; Herr, H.D.; Schaake, J.C. A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrol. Earth Syst. Sci. Discuss.* **2006**, *3*, 1987–2035. [CrossRef]
32. Sikorska-Senoner, A.E.; Quilty, J.M. A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. *Environ. Model. Softw.* **2021**, *143*, 105094. [CrossRef]
33. Tyralis, H.; Papacharalampous, G.; Burnetas, A.; Langousis, A. Hydrological post-processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS. *J. Hydrol.* **2019**, *577*, 123957. [CrossRef]
34. Weerts, A.H.; Winsemius, H.C.; Verkade, J.S. Estimation of predictive hydrological uncertainty using quantile regression: Examples from the national flood forecasting system (England and Wales). *Hydrol. Earth Syst. Sci.* **2011**, *15*, 255–265. [CrossRef]
35. Yan, J.; Liao, G.Y.; Gebremichael, M.; Shedd, R.; Vallee, D.R. Characterizing the uncertainty in river stage forecasts conditional on point forecast values. *Water Resour. Res.* **2014**, *48*, W12509. [CrossRef]
36. Ye, A.; Duan, Q.; Yuan, X.; Wood, E.F.; Schaake, J. Hydrologic post-processing of MOPEX streamflow simulations. *J. Hydrol.* **2014**, *508*, 147–156. [CrossRef]
37. Zhao, L.; Duan, Q.; Schaake, J.; Ye, A.; Xia, J. A hydrologic post-processor for ensemble streamflow predictions. *Adv. Geosci.* **2011**, *29*, 51–59. [CrossRef]
38. Li, W.; Duan, Q.; Miao, C.; Ye, A.; Gong, W.; Di, Z. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *WIREs Water* **2017**, *4*, e1246. [CrossRef]
39. Althoff, D.; Rodrigues, L.N.; Bazame, H.C. Uncertainty quantification for hydrological models based on neural networks: The dropout ensemble. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1051–1067. [CrossRef]
40. Beven, K.; Binley, A. The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process.* **1992**, *6*, 279–298. [CrossRef]
41. Montanari, A.; Koutsoyiannis, D. A blueprint for process-based modeling of uncertain hydrological systems. *Water Resour. Res.* **2012**, *48*, W09555. [CrossRef]
42. Hernández-López, M.R.; Francés, F. Bayesian joint inference of hydrological and generalized error models with the enforcement of total laws. *Hydrol. Earth Syst. Sci. Discuss.* **2017**. [CrossRef]
43. Kuczera, G.; Kavetski, D.; Franks, S.; Thyer, M. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *J. Hydrol.* **2006**, *331*, 161–177. [CrossRef]

44. Vrugt, J.A.; Robinson, B.A. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.* **2007**, *43*, W01411. [CrossRef]
45. Raftery, A.E.; Gneiting, T.; Balabdaoui, F.; Polakowski, M. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather. Rev.* **2005**, *133*, 1155–1174. [CrossRef]
46. Reggiani, P.; Weerts, A.H. A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine. *J. Hydrol.* **2008**, *356*, 56–69. [CrossRef]
47. Todini, E. Paradigmatic changes required in water resources management to benefit from probabilistic forecasts. *Water Secur.* **2018**, *3*, 9–17. [CrossRef]
48. Koutsoyiannis, D.; Montanari, A. Bluecat: A Local Uncertainty Estimator for Deterministic Simulations and Predictions. 2021. Available online: https://doi.org/10.13140/RG.2.2.23863.65445 (accessed on 13 November 2021).
49. Sikorska, A.E.; Montanari, A.; Koutsoyiannis, D. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. *J. Hydrol. Eng.* **2015**, *20*, A4014009. [CrossRef]
50. Robert, C. *The Bayesian Choice*; Springer: New York, NY, USA, 2007. [CrossRef]
51. Gneiting, T. Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **2011**, *106*, 746–762. [CrossRef]
52. Koenker, R.W.; Bassett, G., Jr. Regression quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
53. Koenker, R.W. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.
54. Taylor, J.W. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *J. Forecast.* **2000**, *19*, 299–311. [CrossRef]
55. Tyralis, H.; Papacharalampous, G.; Langousis, A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* **2019**, *11*, 910. [CrossRef]
56. Tyralis, H.; Papacharalampous, G. Boosting algorithms in energy research: A systematic review. *Neural Comput. Appl.* **2021**, *33*, 14101–14117. [CrossRef]
57. Pande, S. Quantile hydrologic model selection and model structure deficiency assessment: 1. Theory. *Water Resour. Res.* **2013**, *49*, 5631–5657. [CrossRef]
58. Pande, S. Quantile hydrologic model selection and model structure deficiency assessment: 2. Applications. *Water Resour. Res.* **2013**, *49*, 5658–5673. [CrossRef]
59. Coron, L.; Thirel, G.; Delaigue, O.; Perrin, C.; Andréassian, V. The suite of lumped GR hydrological models in an R package. *Environ. Model. Softw.* **2017**, *94*, 166–171. [CrossRef]
60. Koenker, R.W. Quantile regression: 40 years on. *Annu. Rev. Econ.* **2017**, *9*, 155–176. [CrossRef]
61. Brehmer, J.R.; Gneiting, T. Scoring interval forecasts: Equal-tailed, shortest, and modal interval. *Bernoulli* **2021**, *27*, 1993–2010. [CrossRef]
62. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [CrossRef]
63. Brehmer, J.R.; Gneiting, T. Properization: Constructing proper scoring rules via Bayes acts. *Ann. Inst. Stat. Math.* **2020**, *72*, 659–673. [CrossRef]
64. Koenker, R.W.; D'Orey, V. Computing regression quantiles. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1987**, *36*, 383–393. [CrossRef]
65. Koenker, R.W.; D'Orey, V. A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1994**, *43*, 410–414. [CrossRef]
66. Koenker, R.W.; Machado, J.A.F. Goodness of fit and related inference processes for quantile regression. *J. Am. Stat. Assoc.* **1999**, *94*, 1296–1310. [CrossRef]
67. Bentzien, S.; Friederichs, P. Decomposition and graphical portrayal of the quantile score. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 1924–1934. [CrossRef]
68. Nascimento, N.O.; Yang, X.; Makhlouf, Z.; Michel, C. GR3J: A daily watershed model with three free parameters. *Hydrol. Sci. J.* **1999**, *44*, 263–277. [CrossRef]
69. Perrin, C.; Michel, C.; Andréassian, V. Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* **2003**, *279*, 275–289. [CrossRef]
70. Pushpalatha, R.; Perrin, C.; Le Moine, N.; Mathevet, T.; Andréassian, V. A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *J. Hydrol.* **2011**, *411*, 66–76. [CrossRef]
71. Mouelhi, S.; Michel, C.; Perrin, C.; Andréassian, V. Stepwise development of a two-parameter monthly water balance model. *J. Hydrol.* **2006**, *318*, 200–214. [CrossRef]
72. Mouelhi, S.; Michel, C.; Perrin, C.; Andréassian, V. Linking stream flow to rainfall at the annual time step: The Manabe bucket model revisited. *J. Hydrol.* **2006**, *328*, 283–296. [CrossRef]
73. Michel, C. *Hydrologie Appliquée aux Petits Bassins Ruraux*; Cemagref: Antony, France, 1991.
74. Addor, N.; Newman, A.J.; Mizukami, N.; Clark, M.P. *Catchment Attributes for Large-Sample Studies*; UCAR/NCAR: Boulder, CO, USA, 2017. Available online: https://doi.org/10.5065/D6G73C3Q (accessed on 13 November 2021).
75. Addor, N.; Newman, A.J.; Mizukami, N.; Clark, M.P. The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 5293–5313. [CrossRef]

76. Newman, A.J.; Sampson, K.; Clark, M.P.; Bock, A.; Viger, R.J.; Blodgett, D. *A Large-Sample Watershed-Scale Hydrometeorological Dataset for the Contiguous USA*; UCAR/NCAR: Boulder, CO, USA, 2014. Available online: https://doi.org/10.5065/D6MW2F4D (accessed on 13 November 2021).

77. Newman, A.J.; Clark, M.P.; Sampson, K.; Wood, A.; Hay, L.E.; Bock, A.; Viger, R.J.; Blodgett, D.; Brekke, L.; Arnold, J.R.; et al. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 209–223. [CrossRef]

78. Newman, A.J.; Mizukami, N.; Clark, M.P.; Wood, A.W.; Nijssen, B.; Nearing, G. Benchmarking of a physically based hydrologic model. *J. Hydrometeorol.* **2017**, *18*, 2215–2225. [CrossRef]

79. Thornton, P.E.; Thornton, M.M.; Mayer, B.W.; Wilhelmi, N.; Wei, Y.; Devarakonda, R.; Cook, R.B. *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2*; ORNL DAAC: Oak Ridge, TN, USA, 2014. Available online: https://doi.org/10.3334/ORNLDAAC/1219 (accessed on 20 January 2016).

80. Oudin, L.; Hervieu, F.; Michel, C.; Perrin, C.; Andréassian, V.; Anctil, F.; Loumagne, C. Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *J. Hydrol.* **2005**, *303*, 290–306. [CrossRef]

81. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]

82. Dunsmore, I.R. A Bayesian approach to calibration. *J. R. Stat. Society Ser. B (Methodol.)* **1968**, *30*, 396–405. [CrossRef]

83. Waldmann, E. Quantile regression: A short story on how and why. *Stat. Model.* **2018**, *18*, 203–218. [CrossRef]

84. Papacharalampous, G.; Tyralis, H. Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. *J. Hydrol.* **2020**, *590*, 125205. [CrossRef]

85. Papacharalampous, G.; Tyralis, H.; Koutsoyiannis, D. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 481–514. [CrossRef]

86. Papacharalampous, G.; Tyralis, H.; Papalexiou, S.M.; Langousis, A.; Khatami, S.; Volpi, E.; Grimaldi, S. Global-scale massive feature extraction from monthly hydroclimatic time series: Statistical characterizations, spatial patterns and hydrological similarity. *Sci. Total Environ.* **2021**, *767*, 144612. [CrossRef] [PubMed]

87. Perrin, C.; Michel, C.; Andréassian, V. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.* **2001**, *242*, 275–301. [CrossRef]

88. Khatami, S.; Peel, M.C.; Peterson, T.J.; Western, A.W. Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. *Water Resour. Res.* **2019**, *55*, 8922–8941. [CrossRef]

89. Khatami, S.; Peterson, T.J.; Peel, M.C.; Western, A.W. Evaluating Catchment Models as Multiple Working Hypotheses: On the Role of Error Metrics, Parameter Sampling, Model Structure, and Data Information Content. 2020. Available online: https://doi.org/10.1002/essoar.10504066.1 (accessed on 13 November 2021).

90. Tyralis, H.; Papacharalampous, G.; Tantanee, S. How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset. *J. Hydrol.* **2019**, *574*, 628–645. [CrossRef]

91. Tyralis, H.; Papacharalampous, G.; Langousis, A.; Papalexiou, S.M. Explanation and probabilistic prediction of hydrological signatures with statistical boosting algorithms. *Remote Sens.* **2021**, *13*, 333. [CrossRef]

92. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021. Available online: https://www.R-project.org/ (accessed on 13 November 2021).

93. Coron, L.; Delaigue, O.; Thirel, G.; Dorchies, D.; Perrin, C.; Michel, C. airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling. R Package Version 1.6.12. 2021. Available online: https://CRAN.R-project.org/package=airGR (accessed on 13 November 2021).

94. Dowle, M.; Srinivasan, A. data.table: Extension of 'data.frame'. R Package Version 1.14.2. 2021. Available online: https://CRAN.R-project.org/package=data.table (accessed on 13 November 2021).

95. Wickham, H.; Hester, J.; Chang, W. devtools: Tools to Make Developing R Packages Easier. R Package Version 2.4.2. 2021. Available online: https://CRAN.R-project.org/package=devtools (accessed on 13 November 2021).

96. Warnes, G.R.; Bolker, B.; Gorjanc, G.; Grothendieck, G.; Korosec, A.; Lumley, T.; MacQueen, D.; Magnusson, A.; Rogers, J. Gdata: Various R Programming Tools for Data Manipulation. R Package Version 2.18.0. 2017. Available online: https://CRAN.R-project.org/package=gdata (accessed on 13 November 2021).

97. Xie, Y. knitr: A Comprehensive Tool for Reproducible Research in R. In *Implementing Reproducible Computational Research*; Stodden, V., Leisch, F., Peng, R.D., Eds.; Chapman and Hall/CRC: London, UK, 2014.

98. Xie, Y. *Dynamic Documents with R and Knitr*, 2nd ed.; Chapman and Hall/CRC: London, UK, 2015.

99. Xie, Y. Knitr: A General-Purpose Package for Dynamic Report Generation in R. R Package Version 1.36. 2021. Available online: https://CRAN.R-project.org/package=knitr (accessed on 13 November 2021).

100. Allaire, J.J.; Xie, Y.; McPherson, J.; Luraschi, J.; Ushey, K.; Atkins, A.; Wickham, H.; Cheng, J.; Chang, W.; Iannone, R. Rmarkdown: Dynamic Documents for R. R Package Version 2.11. 2021. Available online: https://CRAN.R-project.org/package=rmarkdown (accessed on 13 November 2021).

101. Gagolewski, M. Stringi: Character String Processing Facilities. R Package Version 1.7.5. 2021. Available online: https://CRAN.R-project.org/package=stringi (accessed on 13 November 2021).

102. Wickham, H. Tidyverse: Easily Install and Load the 'Tidyverse'. R Package Version 1.3.1. 2021. Available online: https: //CRAN.R-project.org/package=tidyverse (accessed on 13 November 2021).
103. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]