ORIGINAL ARTICLE

CIWEM Chartered Institution of Water and Environmental Management Journal of Flood Risk Management **WILEY**

Comparison of different quantile regression methods to estimate predictive hydrological uncertainty in the Upper Chao Phraya River Basin, Thailand

Suwash C. Acharya[1,2] | **Mukand S. Babel**[2] | **Henrik Madsen**[3] | **Piyamarn Sisomphon**[4] | **Sangam Shrestha**[2]

[1]Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Australia

[2]Water Engineering and Management, Asian Institute of Technology, Khlong Nueng, Thailand

[3]DHI, Hørsholm, Denmark

[4]Hydro-Informatics Institute, Bangkok, Thailand

Correspondence
Suwash C. Acharya, Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Australia.
Email: suwasha@student.unimelb.edu.au, suwashacharya@gmail.com

Abstract

The estimation of predictive uncertainty and its application as a post-processor of hydrological model output, such as water level, can provide additional information useful for short-term hydrological forecasting. In this study, We applied quantile regression models for estimating predictive hydrological uncertainty and used it to derive probabilistic hydrological forecasts. Forecast water levels and associated forecast errors were used as predictor and predictand, respectively, to develop three regression models: (a) linear quantile regression (LQR), (b) weighted LQR and (c) LQR in Gaussian space using Normal Quantile Transformation. These different models for hydrological forecasting were developed for, and applied to, the operational flood forecasting system in the Upper Chao Phraya River, Thailand. The quality of these forecasts in terms of reliability, sharpness and overall skill were assessed using various graphical and numerical verification metrics. Results show that the improvement of forecast in terms of either reliability or sharpness depends upon the configurations used. With comparable overall performance, weighted LQR provided a relatively simple configuration, which can be used for estimating uncertainty in hydrological forecasting.

KEYWORDS

hydrological forecasting, interval score, normal quantile transformation, predictive uncertainty, quantile regression, skill score

1 | INTRODUCTION

Hydrological forecasting is used to predict the future state of a hydrological system to support water management. It is relevant on different temporal scales, ranging from short and medium range (few hours to some days ahead)—such as for flood forecasting and warning and emergency management—to seasonal and long-range forecasting for optimising reservoir operations, water allocation and drought management. Accurate and reliable hydrological forecasts can reduce the impacts from water-related hazards and provide the basis for more effective water resources management.

In this context, information on predictive uncertainty is important for forecast-based decision making (Arnal et al., 2016; Krzysztofowicz, 2002; Ramos, van Andel, &

Pappenberger, 2013; Verkade & Werner, 2011). The probabilistic information on forecasts can assess reliability, resolution and discrimination as measures of forecast quality and hence serves as an essential step for improving the quality of forecasts (Murphy, 1993). Additional information of uncertainty in probabilistic forecasts over deterministic forecasts allows authorities to set risk-based criteria to make rational decisions and also offers potential for additional economic benefits accrued from forecasts (Krzysztofowicz, 2001; Todini, 2017). The use of probabilistic forecasts in risk-based decision making yields higher flood risk reduction, as compared to deterministic forecast-based decision making (Verkade & Werner, 2011).

Uncertainty in hydrological forecasts cannot be eliminated, however, its estimation can help to generate probabilistic forecasts by processing deterministic forecasts (Krzysztofowicz, 2001). Different sources of uncertainties like meteorological information, the initial condition and the model uncertainty are associated with streamflow prediction (Cloke & Pappenberger, 2009). However, the estimation of total predictive uncertainty is a first step towards a better understanding of the uncertainty in forecasts. This total predictive uncertainty can be effectively computed using statistical tool which is further applied to generate probabilistic forecasting for the short and medium ranges.

Todini (2009) defined predictive uncertainty as the 'expression of our assessment of the probability of occurrence of a future (real) event conditional upon all the knowledge available up to the present and the information we were able to acquire through a learning inferential process.'As the definition suggests, uncertainty can be expressed in terms of the probability of exceedance over a threshold, or in terms of a probability distribution conditioned upon hydrological forecasts. The former can be computed from the latter and hence was used to express uncertainty in this study.

Predictive uncertainty can be estimated by developing joint distributions of forecasts and observations (Montanari & Brath, 2004) or by analysing a series of forecast errors. This process usually requires long data series. Since, extreme events such as floods are rare, the estimation of such events is highly uncertain (Cloke & Pappenberger, 2009; Pappenberger et al., 2005; Weerts, Winsemius, & Verkade, 2011). Various methods to estimate predictive uncertainty are available, including quantile regression (QR; Bremnes, 2004; Dogulu, López López, Solomatine, Weerts, & Shrestha, 2015; Koenker & Hallock, 2001; López López, Verkade, Weerts, & Solomatine, 2014; Weerts et al., 2011), Bayesian model averaging (BMA; Raftery, Gneiting, Balabdaoui, & Polakowski, 2005), the hydrologic uncertainty processor (HUP; Krzysztofowicz & Kelly, 2000), the

model conditional processor (MCP; Todini, 2008) and the nonparametric data-based approach (NDA) (Van Steenbergen, Ronsyn, & Willems, 2012), multi-temporal approach of the model conditional processor (MCP-MT; Barbetta, Coccia, Moramarco, Brocca, & Todini, 2017) and uncertainty estimation based on local errors and clustering (UNEEC; Solomatine & Shrestha, 2009). In addition, a summary of Bayesian approaches used in flood forecasting are reviewed in Han and Coulibaly (2017). However, we focused on the application of QR and its configurations in this study.

Quantile Regression, in contrast to Ordinary Least Squares regression, extends the approach of classical regression to estimate conditional quantiles rather than means, which makes it robust to outliers. It can also describe the full distribution as any quantile can be computed. Another important strength of using quantile regression is that QR does not require any assumption based on a prior distribution and can be computed irrespective on distribution of the predictand (Y) and predictor (X) used.

The heteroscedastic and non-linear behaviour of the distribution of forecast errors often seen in hydrological applications should be accounted for in the regression. Apart from using complex non-linear regression, transformation into Gaussian space allows us to use linear regression in Gaussian space, which is then transformed back to the original space (Dogulu et al., 2015; López López et al., 2014; Muthusamy, Godiksen, & Madsen, 2016; Roscoe, Weerts, & Schroevers, 2012; Weerts et al., 2011). This transformation method introduces non-linearity in the regression line in the original space (after the back transformation). This method redistributes the data based on the rank in the Gaussian space, thereby allowing different distributions of forecast errors in different ranges of forecast values. The back-transformation may not be straight-forward if the quantile to be back-transformed lies outside the range of calibration data and could introduce more uncertainty. This is particularly important if we are considering large events, where, decisions are to be made based on the forecasts. This issue has been treated using linear extrapolation at both extreme ends by López López et al. (2014) and Weerts et al. (2011), however, the performance of such a method was not evaluated in these studies due to lack of extreme events in the validation dataset.

In this study, QR was applied to forecast data using a forecast system of the Chao Phraya River Basin (CPRB) in Thailand. Major floods in CPRB have occurred several times in the past: in 1942, 1978, 1983, 1995, 1996, 2002, 2006, 2010 and 2011. Despite different structural and non-structural measures in place, the flooding problem in CPRB still exists. The recent flood in 2011, which

caused devastating damage of property, has emphasised the importance of a forecasting system and decision-making based on the system. A real-time flood forecasting and management decision support system based on MIKE, customised by DHI, was set up in order to compensate for the lack of reliable flood forecast data during the monsoons (Sisomphon, Boonya-aroonnet, & Chonwattana, 2013). However, the question of the reliability of the forecasting system continues to exist as the forecasting system is currently not able to provide an estimate of the uncertainty of the forecasts. This study, driven by a potential of shift from deterministic forecasting towards probabilistic forecasting, investigated methods based on QR for the provision of predictive hydrological uncertainty associated with the flow forecasting system. This study applied and compared three different configurations of QR: (a) Linear quantile regression (LQR), (b) Weighted LQR that gives larger weights to higher water levels in the regression and (c) LQR in Gaussian space using Normal Quantile Transformation. We evaluated these QR models with the aim of investigating whether the non-linearity introduced in the regression lines in the original space by applying QR in Gaussian space provide a good representation of the actual error and increases the performance of the QR model in comparison to other configurations. In addition, we explored whether a simpler weighted QR configuration could perform on par or better than the method involving transformation of the data. These methods were compared using a range of probabilistic verification metrics.

2 | STUDY AREA

The area for this study is CPRB in Thailand. It is located entirely within Thailand and covers an area of 157,925 km^2. The river drains from north to south and finally into the Gulf of Thailand (Figure 1). The basin consists of four large tributaries, namely, Ping, Wang, Yom and Nan, which merge into the main Chao Phraya River. Chao Phraya River Basin has a monsoon dominated weather regime with a rainy season lasting from May to October. The annual average rainfall is 1,153 mm and the annual averaged runoff is 3,762 MCM. About 90% of the annual rainfall occurs during the rainy season, causing high discharge and heavy floods in the area.

The present study focuses on the Upper CPRB (Figure 1). The Upper Chao Phraya rivers—Ping, Wang, Yom and Nan—flow into the main Chao Phraya river at Nakhon Sawan province. The upper CPRB is a hilly area and covers approximately 109,973 km^2. The discharge of Ping, Wang and Nan rivers are mainly controlled by Bhumibol Dam, Kiew Lom Dam and Sirikit Dam, respectively.

The flood forecasting system, which is currently being operated by the Hydro-Informatics Institute was used. This real-time flood forecasting system is composed of a weather forecasting system, catchment hydrological model, river hydrodynamic model and a data assimilation system for real time model updating. The weather research and forecasting (WRF) model is used to forecast rainfall for 7 days ahead. The input data for WRF is derived from the NCEP Global Forecasting System. The hydrological models are implemented using the NAM module of the MIKE Powered by DHI MIKE 11 software. NAM is a lumped, conceptual rainfall-runoff model. The hydrodynamic models are implemented using the MIKE 11 river modelling system. It is a one-dimensional hydrodynamic software package including a full solution of the St. Venant equations. The MIKE 11 model calculates water levels and discharges along the modelled river stretches. As part of the forecasting system, data assimilation is carried out to update and adjust the MIKE 11 model to available observations. The total simulation period of a forecast run is 14 days, consisting of 7 days of hindcast (with data assimilation) and 7 days of forecast. More information about the forecast system can be found in Sisomphon et al. (2013).

The operationally available forecasted and observed water level data for three stations: NAN007, NAN008 and PIN004 were used in this study. These stations are located at the lower part of the Upper CPRB as shown in Figure 1. For this study, the available data was split into calibration (09/2013–09/2014) and validation datasets (10/2014–12/2014).

3 | METHODOLOGY

3.1 | Quantile regression

Quantile regression is a regression of conditional quantiles. For a given random variable Y, the τ'th quantile is the value y for which the probability of occurrence of a quantity below this value is equal to τ, that is, $P(Y < y_\tau) = \tau$. The interval between quantiles gives the confidence interval. For example, 25 and 75% quantiles give 50% confidence interval, whereas 5 and 95% quantiles give 90% confidence interval. In terms of distribution, the cumulative distribution function for the random variable Y can be written as $F(y) = P(Y < y)$, and the quantile function is the inverse of this function, $Q(\tau) = F^{-1}(\tau)$.

QR minimises the absolute sum of deviations from the regression line. For two variables X and Y whose
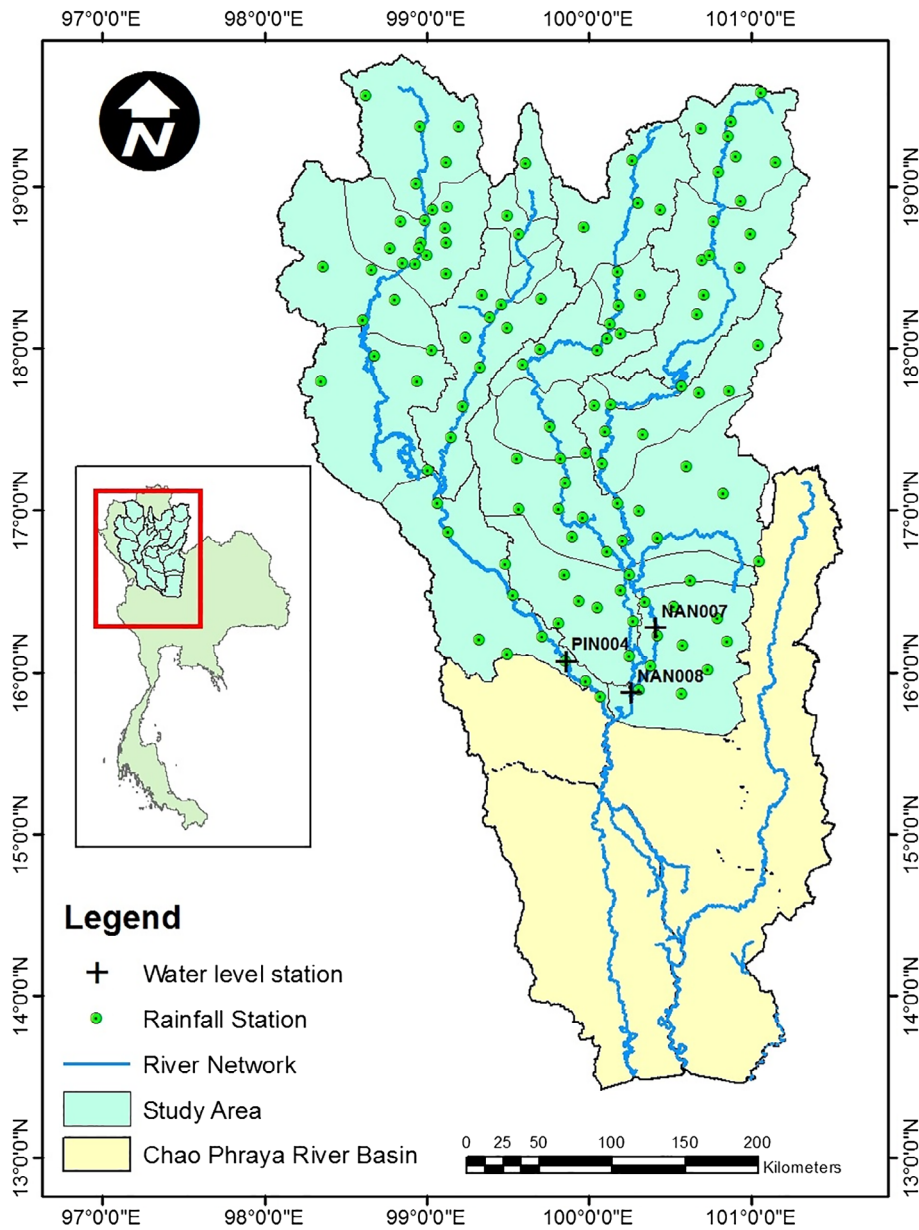
4 of 16 | WILEY—**CIWEM** Chartered Institution of Water and Environmental Management Journal of Flood Risk Management

ACHARYA ET AL.

relation is to be established using quantile regression at a quantile τ, the conditional quantile regression can be expressed as a minimization problem as:

$$\min \sum_{i=1}^{n} \rho_\tau (y_i - \hat{y}_\tau (x_i, \beta_\tau)) \qquad (1)$$

where (x_i, y_i), $i = 1,2,...,n$ is the sample of (X,Y), $\hat{y}_\tau$ is the estimate of the quantile of Y, β_τ represents the parameters of the regression line and $\rho_\tau(.)$ is an asymmetric loss function given by:

$$\rho_\tau(y) = \begin{cases} (\tau-1)\,y, & y \leq 0 \\ \tau\,y, & y > 0 \end{cases} \qquad (2)$$

From the available series of water level forecasts (X) and observations (O), the forecast error at a given lead time of interest, *t*, can be computed as:

$$Y(t) = O(t) - X(t) \qquad (3)$$

It was assumed that, for a given lead time, the forecast error changes (mostly increases) with the forecast value and we were interested in determining the quantiles of the error distribution as a function of the forecast using QR.

The regression parameters for each quantiles are independent of each other and hence may result in the crossing of quantiles. Therefore, a fixed error model described in Weerts et al. (2011) was applied below the

TABLE 1 Various configurations of quantile regression used in the study

Name	Description
LQR	Linear quantile regression in original space
LQR:WT	Linear quantile regression in original space with higher weight applied to higher water level
LQR:NQT	Linear quantile regression in Gaussian space using NQT

crossing point in this study as the crossing usually occurs at lower forecast values.

The different configurations of QR used in this study are: LQR (López López et al., 2014), LQR:NQT (López López et al., 2014; Roscoe et al., 2012; Weerts et al., 2011), and a modification of LQR focusing on higher flows LQR:WT, which are presented in Table 1.

3.1.1 | LQR in original space

This is the simplest and most direct approach of QR using data in original space.

The linear regression equation is described by:

$$Y(t) = \beta_0 + \beta_1 X(t) + \varepsilon \tag{4}$$

where β_0 and β_1 are the intercept and slope of the linear regression line and ε is the residual.

The minimization problem of quantile regression applies the loss function to the residual ε for each quantiles to find the optimum parameters, that is, defined as follows.

$$\min \sum_{i=1}^{n} \rho_\tau \left(Y_i - \left(\beta_{0,\tau} + \beta_{1,\tau} * X_i \right) \right) \tag{5}$$

3.1.2 | LQR in original space with weights (LQR:WT)

Although the estimation of uncertainty is targeted especially for higher flows in the flood forecasting system, few data are available in that flow range. Most of the data are concentrated at lower water levels. This uneven distribution of data could lead to biased regression models when used in the original space. To address this, weighted linear quantile regression is used in the study. This configuration is different from LQR only in the application of weights.

The weighing can be applied by ranking the data on the basis of the forecast water level and then providing the weight based on the rank of the data. In this study, the weight was applied such that the highest ranked data was given the highest weight (= 1), that is,

$$w_i = r_i / N, \tag{6}$$

where r_i is the rank of ith forecast water level data (from lowest to highest), N is the total number of data.

3.1.3 | LQR in Gaussian space using NQT (LQR:NQT)

A transformation such as normal quantile transform (NQT) can be applied to the forecast and error series to make the distributions Gaussian. The NQT transformation was applied to both $Y(t)$ and $X(t)$.

$$Y_T(t) = Q^{-1}[F(Y(t))] \tag{7}$$

$$X_T(t) = Q^{-1}[F(X(t))] \tag{8}$$

where $F(.)$ is the Weibull plotting position and Q^{-1} is the inverse of the standard normal distribution.

Quantile regression was applied thereafter solving the minimization:

$$\min \sum_{i=1}^{n} \rho_\tau \left(Y_{T,i} - \left(\beta_{0,\tau} + \beta_{1,\tau} * X_{T,i} \right) \right) \tag{9}$$

After the analysis in Gaussian space, the variables are back transformed into original space.

3.2 | Forecast evaluation

We compared the performances of the QR models using different metrics. The aspects of forecast evaluation we considered were reliability, sharpness and skill. Reliability is a measure of statistical consistency between the forecast probability and the observed frequency of the events. Sharpness refers to the spread of forecast distribution and is solely a function of forecast. Skill measures the quality of forecast by considering both reliability and sharpness. The verification metrics were chosen based on the type of probabilistic forecast, that is, full continuous forecast probability distribution and central credible interval (CCI) forecasts (Gneiting & Raftery, 2007; Wilks, 2011). The generation of full continuous probability

distributions was done by computing a large number of quantiles over [0,1], whereas some quantiles (0.25 and 0.75 for 50% CI and 0.05 and 0.95 for 90% CI) were selected to generate CCI forecasts. Reliability, continuous ranked probability score (CRPS) and its skill score (CRPSS) were computed as measures of full continuous forecast probability distribution, while prediction interval coverage probability (PICP), mean prediction interval (MPI) and interval score were computed for CCI.

3.2.1 | Reliability diagram

The reliability diagram is an approach used to graphically represent the performance of probability forecasts expressed in terms of quantiles. A reliability diagram consists of a plot of observed relative frequency as a function of forecast probability.

For n pairs of observed O and forecast X, the forecast non-exceedance probability is τ (equal to the quantile of consideration) and the observed relative frequency of non-exceedance is given by f_τ:

$$f_\tau = \frac{1}{n}\sum_{i=1}^{n} 1\{O_i \leq X_\tau\} \tag{10}$$

where 1{.} represents the value of 1 when the inequality inside the bracket is satisfied and 0 otherwise.

For perfect reliability, the forecast non-exceedance probability and observed relative frequency are equal and the reliability line follows the 1:1 diagonal line.

3.2.2 | Continuous ranked probability score

The ranked probability score is essentially an extension of the Brier score to multiple categories. It is used to access the overall forecast performance of probabilistic forecasts. The ranked probability score is the sum of squared differences between the components of the cumulative forecast and observation vectors, and it is given by

$$RPS = \frac{1}{J-1}\sum_{m=1}^{J}\left[\left(\sum_{j=1}^{m} x_j\right) - \left(\sum_{j=1}^{m} o_j\right)\right]^2 \tag{11}$$

where J is the number of events being forecasted, x is the forecast and o is the observation.

Similar to the Brier score, RPS = 0 represents a perfect forecast and a higher RPS reflects a poorer forecast.

For continuous predictands, the RPS is extended to CRPS (Wilks, 2011)

$$CRPS = \int_{-\infty}^{\infty} [F(x) - F_o(x)]^2 dx \tag{12}$$

where $F(x)$ is cumulative distribution function of probabilistic forecast and $F_o(x) = \begin{cases} 0, & x < o \\ 1, & x \geq o \end{cases}$.

3.2.3 | Skill score

Forecast skill measures the quality of a set of any forecast with respect to a reference or benchmark forecast. It is represented by skill score, which gives the percentage improvement of the forecast compared to its reference forecast. For any score S, the corresponding skill score can be calculated as:

$$SS = \frac{S - S_{ref}}{S_{perf} - S_{ref}} \tag{13}$$

where S_{ref} is the score of the reference forecast which corresponds to threshold forecast over which improvement is computed and S_{perf} is the score of the perfect forecast which is equal to the best score that could be achieved. In present study, S = CRPS; S_{ref} = CRPS$_{ref}$, the CRPS score for a reference forecast; S_{perf} = 0, which is maximum value of CRPS score that can be attained for a perfect forecast.

The reference forecasts used for computing skill scores may be calculated by using a reference lumped hydrological model with persistent or climatological meteorological forcing, or may directly be obtained from climatologic or persistent observed discharge (Pappenberger et al., 2015). In this study, a probabilistic distribution based on climatologic observed discharge was used as the reference forecast.

3.2.4 | PICP and MPI

PICP and MPI (Shrestha & Solomatine, 2006) are measures of reliability and sharpness based on the prediction interval. Thordarson et al. (2012) used the same metrics for evaluating probabilistic forecasts of sewer flow. The reliability plot described above shows the reliability in terms of forecast quantiles. However, the forecast uncertainties are often expressed in centred intervals. For a given percentage of centred interval β, the interval with

the quantiles are $[\alpha/2, 1 - \alpha/2]$ where $\alpha = 1 - \beta$. The reliability diagram gives the theoretical quantiles and corresponding observed quantiles, whereas PICP gives the proportion of the data lying in the centred interval. MPI is the average width of the forecast interval. These measures together can be used to compare different methods and their performance for both reliability and sharpness.

For a given model, the upper and lower boundary for the forecast interval are defined as $x_{i,\beta}^u$ and $x_{i,\beta}^l$, respectively, for a given confidence interval β, at forecast time i. The observation at that time is denoted by O_i. Then the hit and miss of the observed value within the given forecast interval can be given by a binary indicator variable:

$$h_{i,\beta} = 1\left\{x_{i,\beta}^u \geq O_i \geq x_{i,\beta}^l\right\} \quad (14)$$

where $1\{.\}$ represents the value of 1 when the inequality inside the bracket is satisfied and 0 otherwise, and $\alpha = (1 - \beta)$ so that the upper and lower quantiles of the prediction interval are $u = 1 - \alpha/2$ and $l = \alpha/2$.

PICP is the fraction of observations occurring within the given interval determined by the expected value of $h_{i,\beta}$ over the range of n forecasts and observations:

$$\text{PICP}_\beta = E\left[h_{i,\beta}\right] = \frac{1}{n}\sum_{i=1}^{n} h_{i,\beta}. \quad (15)$$

MPI for the confidence interval β, expressed in the unit of the forecasted variable, is given by:

$$\text{MPI}_\beta = \frac{1}{n}\sum_{i=1}^{n}\left(x_{i,\beta}^u - x_{i,\beta}^l\right). \quad (16)$$

Better reliability is obtained for PICP_β values closer to the considered confidence interval β. A smaller value of MPI implies a sharper forecast. If the PICP is smaller than the corresponding confidence interval, the PICP can be improved at the expense of an increase in MPI, and vice versa.

3.2.5 | Interval score

The interval score can be expressed as a combination of reliability and sharpness as shown in Gneiting and Raftery (2007) and Thordarson et al. (2012). The interval score is given by:
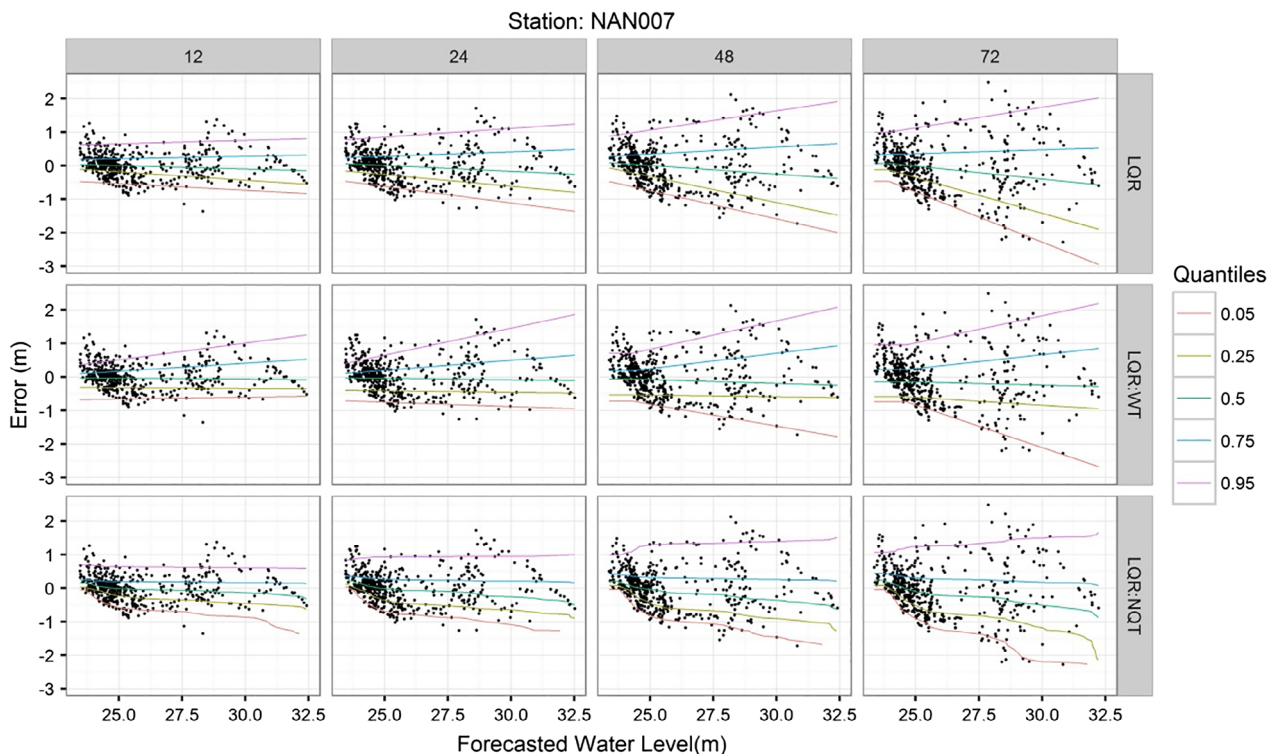

FIGURE 2 Forecast error versus forecast water level with regression lines of various quantiles for NAN007 station. QR configurations are shown in rows and lead time (in hr) is in columns

$$IS_{i,\beta} = \begin{cases} \left(x_{i,\beta}^u - x_{i,\beta}^l\right) + \dfrac{2}{\alpha}\left(x_{i,\beta}^l - O_i\right), & O_i < x_{i,\beta}^l \\[2ex] \left(x_{i,\beta}^u - x_{i,\beta}^l\right) + \dfrac{2}{\alpha}\left(O_i - x_{i,\beta}^u\right), & O_i > x_{i,\beta}^u \end{cases}. \quad (17)$$

For a series of n forecasts and observations, the score is averaged over the whole period to obtain an average interval score:

$$\bar{IS}_\beta = MPI_\beta + \frac{2}{\alpha * n}\sum_{i=1}^{n}\left[\left(x_{i,\beta}^l - O_i\right)1\left\{O_i < x_{i,\beta}^l\right\}\right.$$
$$\left. + \left(O_i - x_{i,\beta}^u\right)1\left\{O_i > x_{i,\beta}^u\right\}\right] \quad (18)$$

Equation (18) shows that the interval score is the sum of MPI and a penalization term for the observations outside the prediction interval range. If the interval gets wider, MPI will increase, which increases the score, on the other hand a smaller MPI will lead to a penalization of the observations that lies outside the prediction interval and thereby increases the overall score. The interval score can also be expressed in terms of reliability and sharpness as:

$$\bar{IS}_\beta = MPI_\beta + \frac{2}{\alpha * n}\sum_{i=1}^{n}\left(1 - h_{i,\beta}\right)\left(\min\left|O_i - \left[x_{i,\beta}^l,\ x_{i,\beta}^u\right]\right|\right) \quad (19)$$

where the second summation term accounts for proportion of observed data lying outside the described prediction interval and the minimum distance from prediction interval (either upper or lower limit) to those observed value.

4 | RESULTS AND DISCUSSION

4.1 | Error models

The QR models generated for the three different stations NAN007, NAN008 and PIN004 are shown in Figures 2–4 respectively. The plots consist of scatter of forecasted water levels and corresponding forecast errors with regression lines obtained from QR. The quantiles shown in the plots are 0.05, 0.25, 0.5, 0.75 and 0.95. However, quantiles were computed with intervals of 0.05 over the [0, 1] domain. The study was conducted for lead time of each hours of forecast up to 168 hr, however, in this paper we presented the results up to a lead time of 72 hr.

The regression model, in principle, could be fitted either on forecast values or forecast errors. Although this investigation is not presented in detail, we found that the regression based on forecast values yielded a narrow CI at higher water level regime and provided unreliable estimates. This effect was particularly mentioned in Weerts et al. (2011) and was also observed in López López et al. (2014) yet not described in detail. On using NQT, the higher water levels in forecasts and observed data in
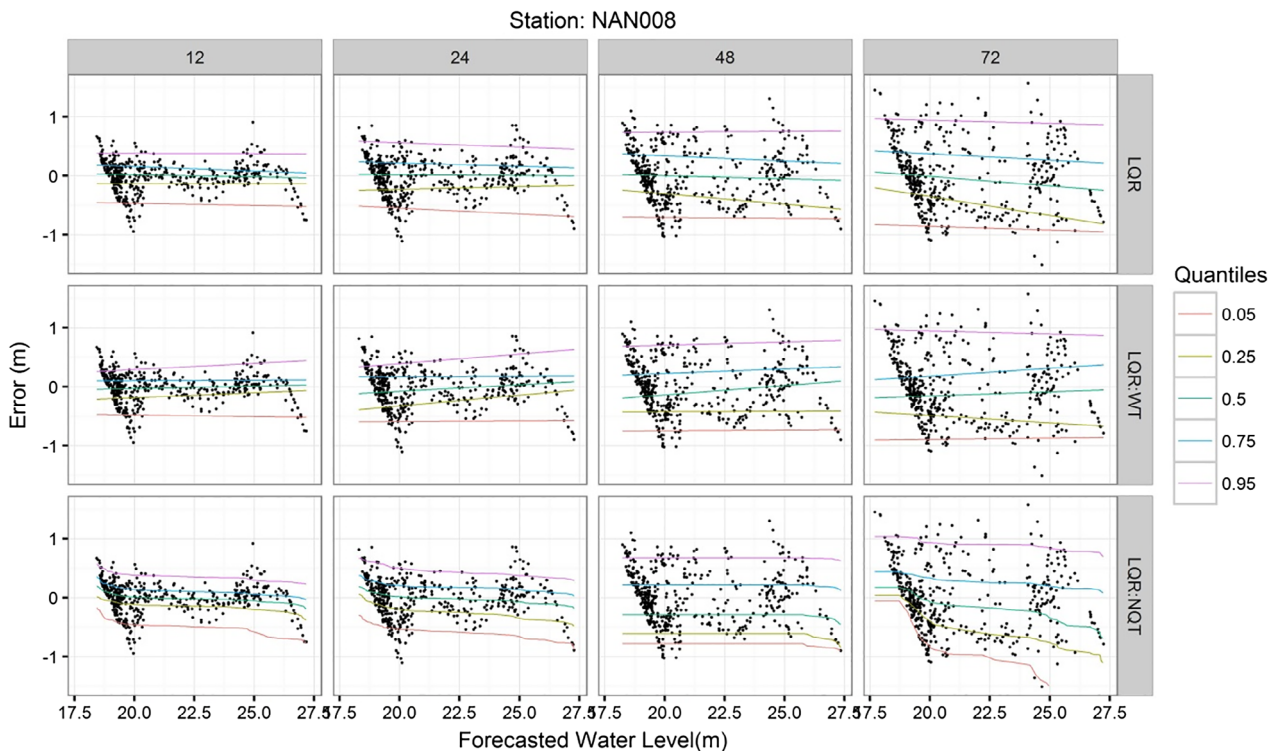

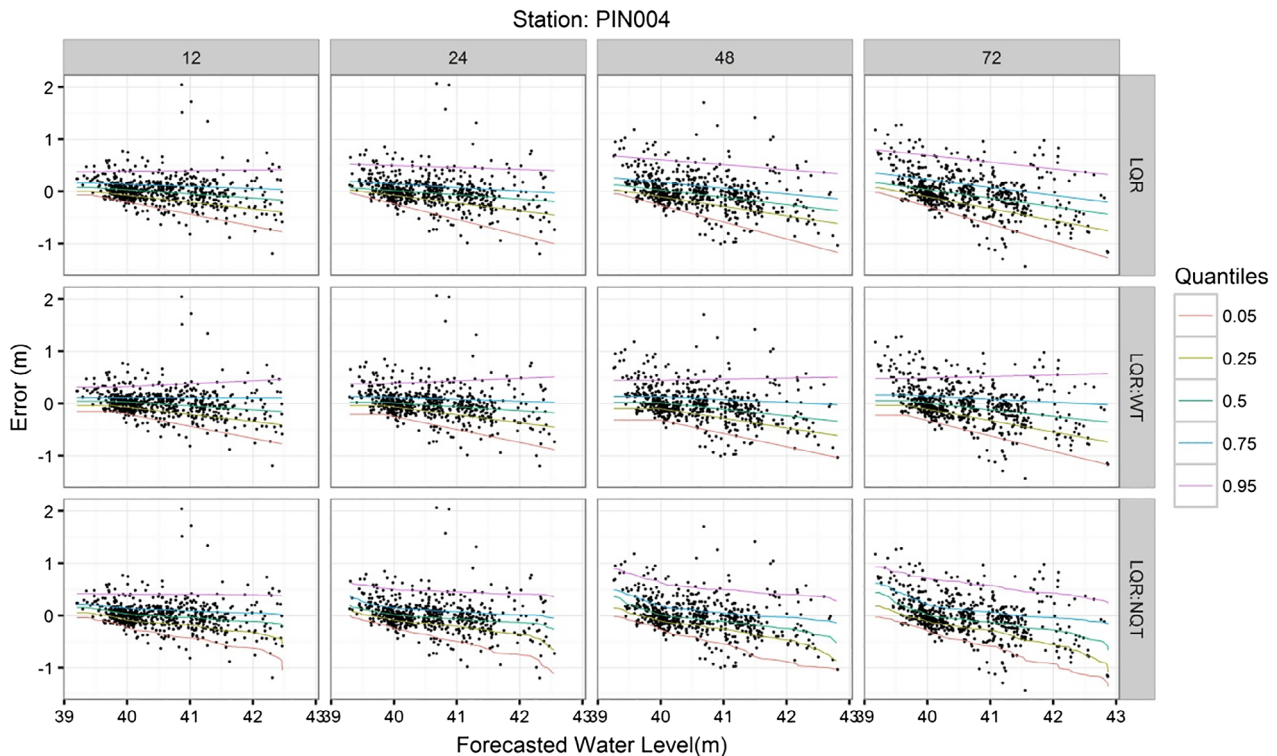**FIGURE 3** Same as Figure 2 for NAN008 station

FIGURE 4 Same as Figure 2 for PIN004 station

transformed space are close to each other irrespective of their magnitude. This will result in convergence of regression lines at that domain. This effect was not observed in LQR and LQR:WT, which does not involve any transformations.

Each station in the study showed different patterns of forecast errors. For station NAN007, the errors increased with the increase in water level forecast with a tendency towards negative errors for larger lead times. Station NAN008 showed no clear pattern of the errors, which are almost uniformly distributed over the full range of the forecasted water levels. The error at station PIN004 increased for increasing water level and shows a biased forecast with negative errors implying over-forecasting at higher flow ranges. For all the stations, the uncertainty tended to increase with lead time which was expected. Similar behaviour of uncertainty was observed in the previous studies as well (López López et al., 2014; Weerts et al., 2011). Despite increasing uncertainty, the patterns of error models were similar across lead times. It is worth noting that the model at different lead times include same set of observation values and ultimately leads to similar pattern.

Configurations LQR and LQR:WT provide linear relationships and are hence quite similar. However, the regression lines in LQR:WT were more influenced by the pattern of the errors at the higher water levels. Due to influence of higher values, crossing of quantiles in LQR:

WT occurred more at higher values than it did in LQR, which was observed at stations NAN007 and PIN004.

Figure 5 shows the regression lines in the Gaussian space. It explains the non-linearity introduced in the LQR:NQT configuration. The linear regression in the Gaussian space from Figure 5 was transformed back to the original domain and is compared to the other two configurations in Figures 2–4. The overall regression line seemed to follow the trend of the data. However, the nonlinear line was affected by the data at a given location. The LQR:NQT regression line seemed to be more horizontal where the data were less dense. This is due to the re-distribution of data in the Gaussian space which is based on the Weibull plotting position. The tendency to under-forecast at NAN007 for lower flows was better captured by LQR:NQT for lower quantiles than higher quantiles. This is due to the application of linear quantile regression, which overlooks local trends and follows the general trend of the data.

The fixed error model which is observed at the lower regime of level does not affect the performance of higher levels. Such fixed error model is clearly visible in lower flow regime. Such error model is sometimes dominant at transformed domain in LQR:NQT, however, it is not even visible when transformed back. For probabilistic flood forecasting, where only large levels are involved, any additional models describing the lower level regimes could be ignored. However, in a general-purpose
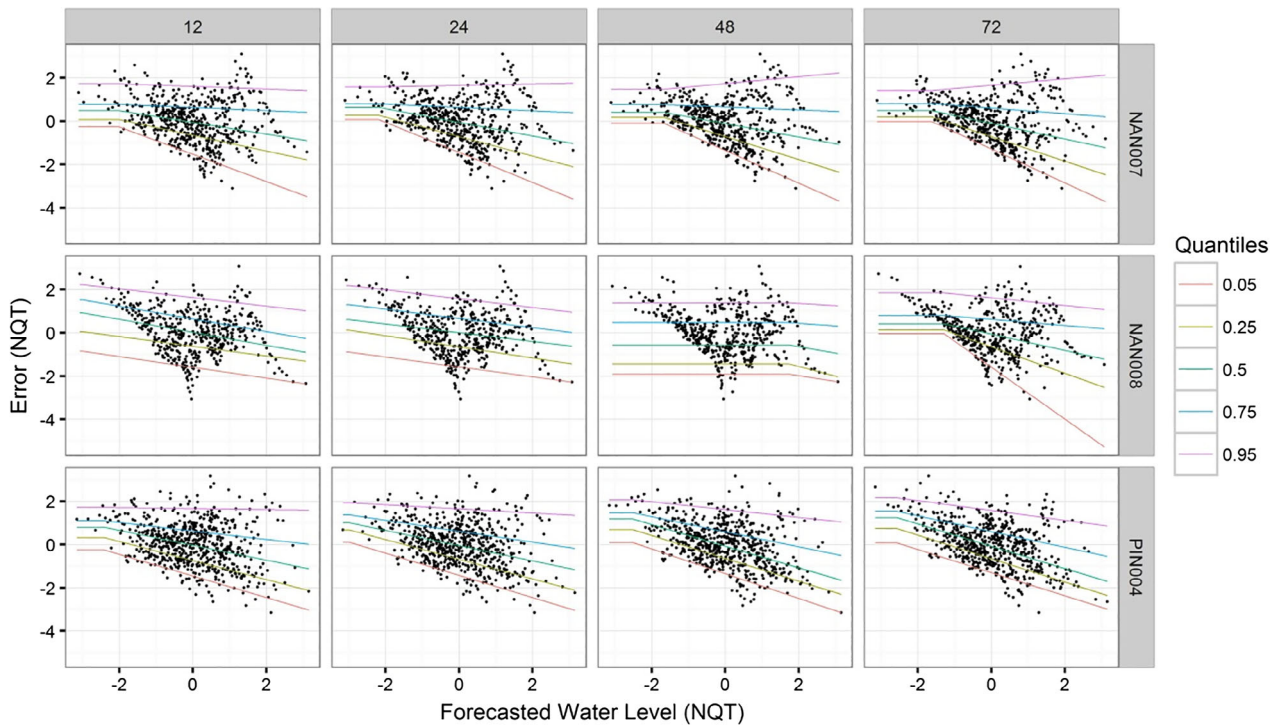
FIGURE 5 Forecast error versus forecast water level with regression lines of various quantiles at different stations in Gaussian space. Stations are shown in rows and lead time (in hr) is in columns
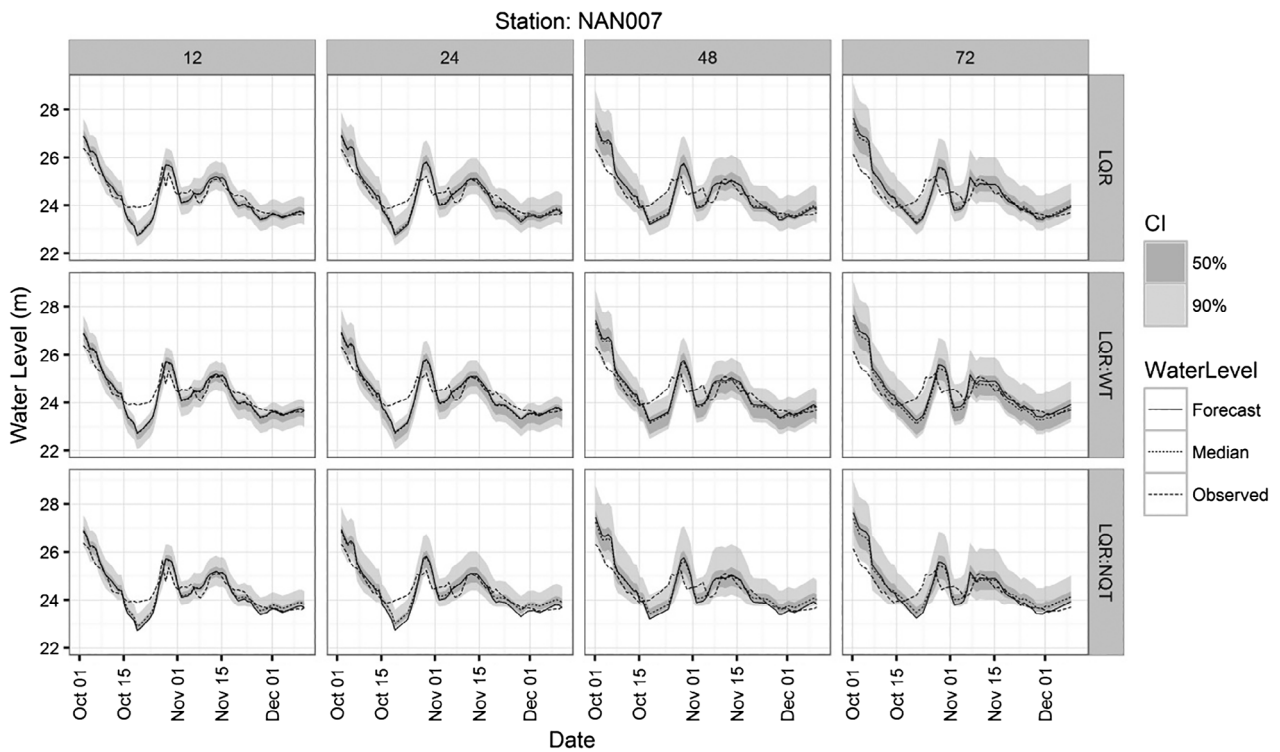

FIGURE 6 Hydrograph at NAN007 station for October–December 2014. Configurations are in rows while lead time (in hr) is in columns. Grey bands show confidence interval of 50 and 90%, and lines show forecasted and observed water levels, and the median of the probabilistic forecast
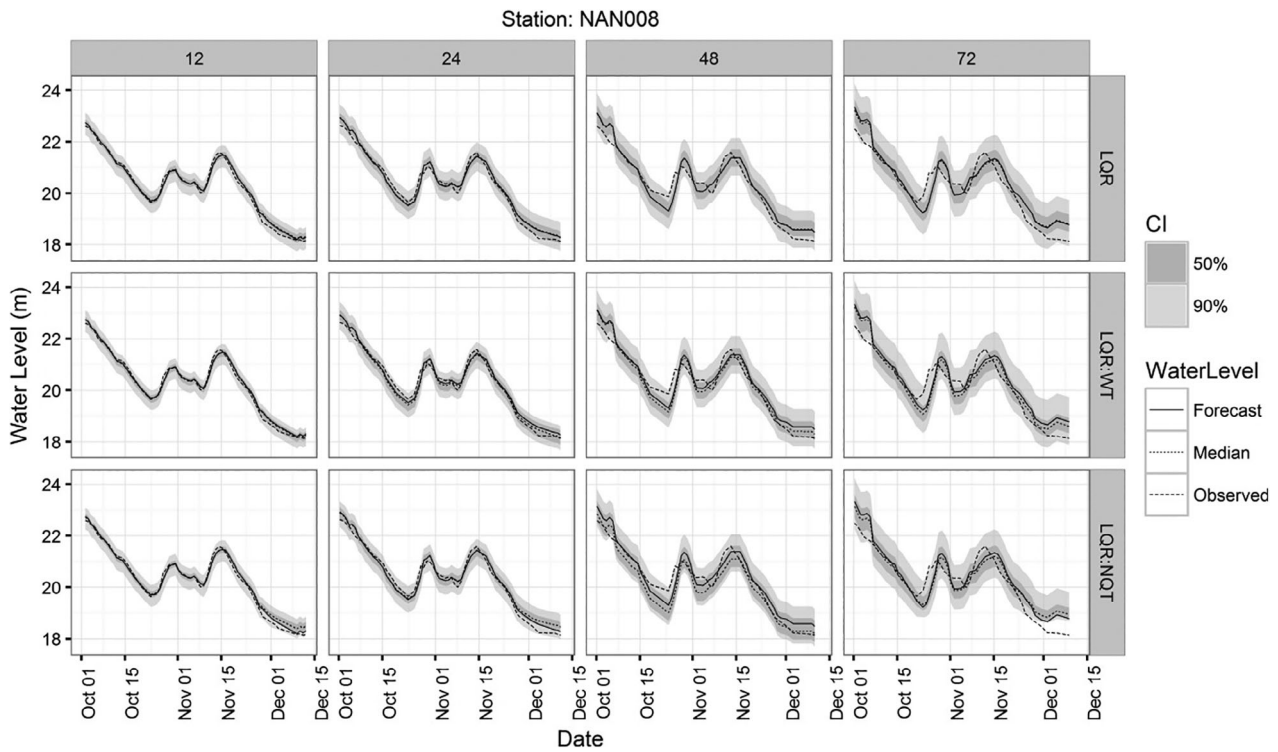
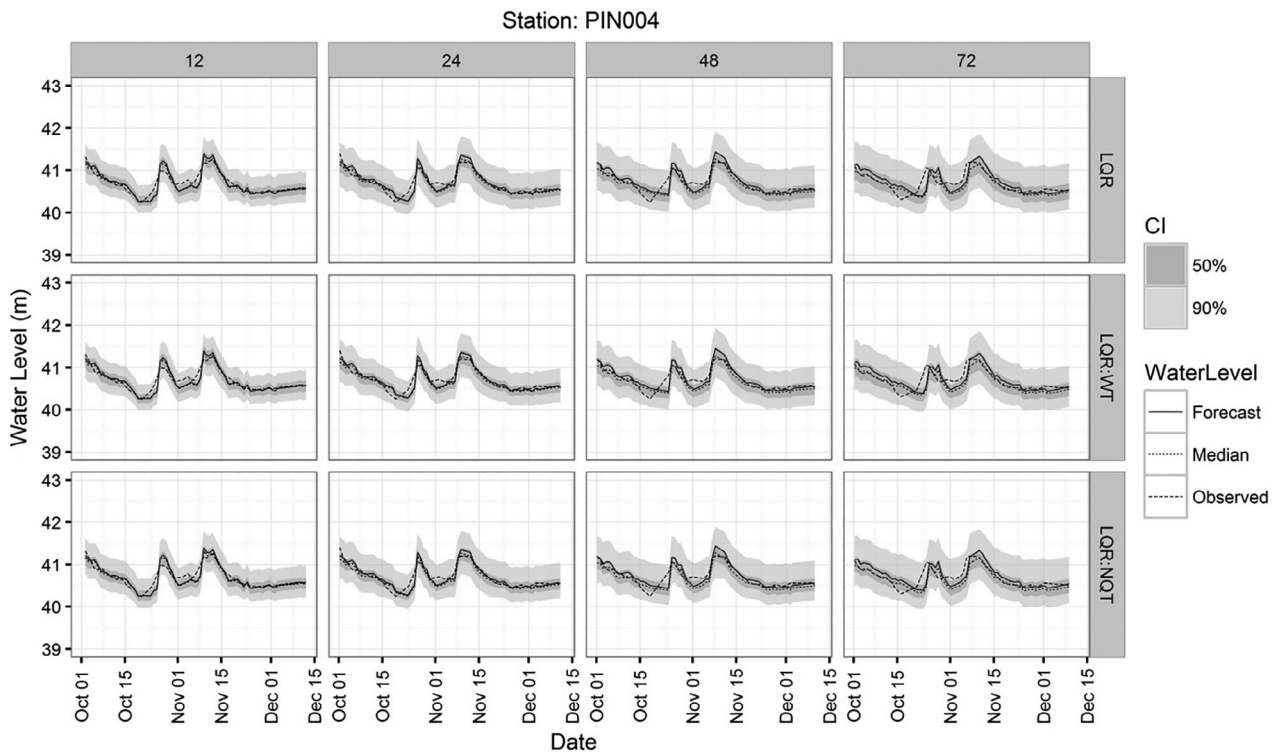**FIGURE 7** Same as Figure 6 for NAN008 station

FIGURE 8 Same as Figure 6 for PIN004 station

hydrological forecasting system, the lower level regimes could be described. Such models may either be a simple fixed error model (Weerts et al., 2011), non-crossing quantiles or stepwise regression (López López et al., 2014).

4.2 | Forecast hydrographs

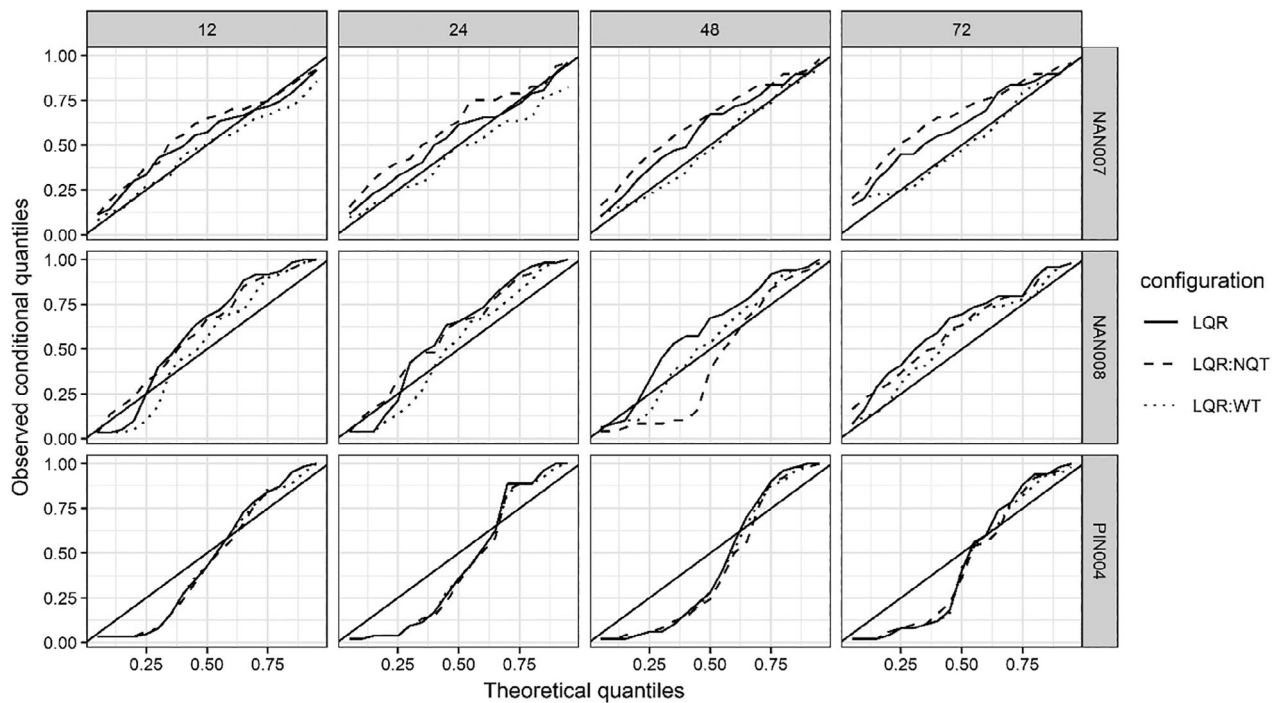The quantile regression models were used to generate probabilistic forecasts for the period October 2014 to

12 of 16 | WILEY—**CIWEM** Chartered Institution of Water and Environmental Management Journal of Flood Risk Management

ACHARYA ET AL.

FIGURE 9 Reliability Q–Q plot for various stations (in rows) at different lead times (in columns, in hr)

TABLE 2 Comparison of MPI for different configurations at different confidence intervals

Station and lead time (hr)	50% CI			90% CI		
	LQR	LQR:WT	LQR:NQT	LQR	LQR:WT	LQR:NQT
NAN007						
12	0.363	0.462	**0.340**	1.160	1.136	**1.016**
24	0.502	0.590	**0.463**	1.408	1.332	**1.278**
48	0.598	0.777	**0.566**	1.708	1.602	**1.597**
72	0.649	0.865	**0.617**	1.817	1.871	**1.676**
NAN008						
12	0.275	0.279	**0.260**	0.852	**0.785**	0.831
24	0.429	0.470	**0.398**	1.111	**1.002**	1.018
48	0.656	**0.652**	0.827	1.457	1.457	**1.454**
72	0.757	**0.704**	**0.704**	1.796	1.822	**1.680**
PIN004						
12	0.280	0.281	**0.267**	0.772	**0.744**	0.786
24	0.281	0.268	**0.266**	0.929	**0.857**	0.892
48	0.330	0.327	**0.297**	1.046	**0.974**	0.996
72	0.391	0.370	**0.343**	1.141	**1.064**	1.128

Note: The numbers in the bold indicate best value among the configurations.

December 2014 for the three stations NAN007, NAN008 and PIN004. The probabilistic forecasts are represented in the form of hydrographs in the plots in Figures 6–8.

The three stations showed different behaviour, which can also be interpreted from the regression models explained in the previous section. For NAN007 and PIN004, the width of the uncertainty bands increased with increased water level as well as with the increase in the lead time, showing forecasts to be more uncertain at higher water levels and larger lead times. However,

NAN008 did not show an increase in uncertainty with increasing water level as seen in the error model (Figure 3). The error model of station PIN004 gave smaller forecast intervals as the error was lower and the uncertainty increased only slightly with the forecast water level.

PIN004 showed biased forecasts (over-forecasting) with the median lying below the forecasted water level for most of the flow regime. For station NAN007, for the lowest flow range, the forecasts were much lower than the observed (under-forecasting). This trend was seen in the calibration dataset also but all three configurations failed to represent the trend of errors at lower flows (Figure 2).

4.3 | Forecast evaluation

Figure 9 shows the reliability Q-Q plots to evaluate reliability of the forecasts. It can be stated that the forecasts are more reliable for stations NAN007 and NAN008 as the lines are closer to the diagonal. The reliability of forecasts for all three QR configurations is similar, with the configuration LQR:WT producing slightly more reliable forecasts compared to the others. For station PIN004, the reliability plot shows S-shaped curve implying that the forecast at this station was under-confident. This can also be seen in the hydrograph as almost all the data lies within the prediction interval indicating that the forecast intervals are wider than required. The MPI of the validation data for 50% CI and 90% CI for each stations and configurations are shown in Table 2 to ascertain the sharpness of the forecast. For 50% CI, LQR:NQT shows smaller confidence band across all lead times for NAN007 and PIN004 stations and at certain lead times for NAN008 station. For 90% CI, the confidence band is smaller in LQR:NQT for NAN007 station, LQR:WT for PIN004 station and mixed results for NAN008 station.

The CRPS requires full distribution of forecasts. Since the forecast of quantiles is less likely to be symmetrical and follow the normal distribution, the empirical cumulative distribution function of forecasts was used in the estimation of CRPS. The climatological CRPS was also calculated and the skill scores for the forecast were computed using Equation (13). Results are shown in Figure 10.

The LQR:NQT configuration showed higher skills as compared to other configurations, for station NAN007. For stations NAN008 and PIN004, the skill scores were close to each other for all configurations. In general, the skill of the forecast decreases with the increase in lead time. It was seen that the performance evaluation using interval score was similar to CRPSS. Also, the quality of
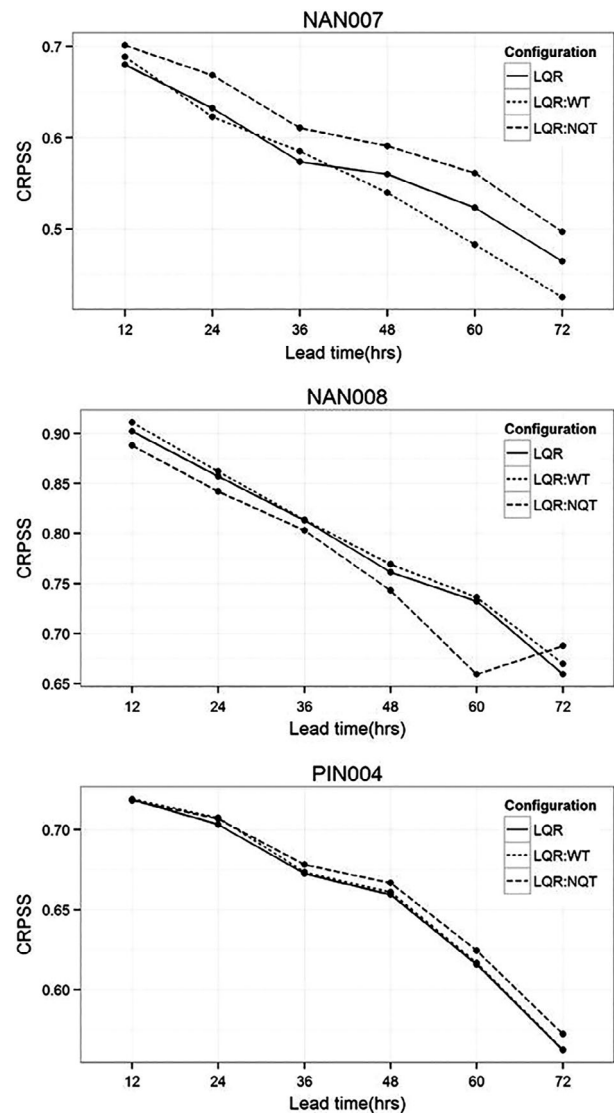

FIGURE 10 Continuous Ranked Probability Skill Score (CRPSS) for different configurations at various lead times at stations NAN007, NAN008 and PIN004

probabilistic forecast in terms of skill depended largely on the quality of deterministic forecasts. The station NAN008 whose forecast error was smaller showed better CRPSS (0.9 at 12 hr to 0.67 at 72 hr) as compared to other stations.

Figure 11 shows the PICP-MPI plots as well as the interval scores. The scores were computed for confidence intervals of 50, 80 and 90%. For NAN007 station, the interval score was minimum for the LQR:NQT configuration, which had, in general, better MPI and almost the same PICP as the other configurations. For NAN008, the interval scores are almost equal at lower lead time but for higher lead time, scores are higher for LQR:NQT which can be explained with lesser reliability from the PICP-MPI diagram. For PIN004, the interval scores were better
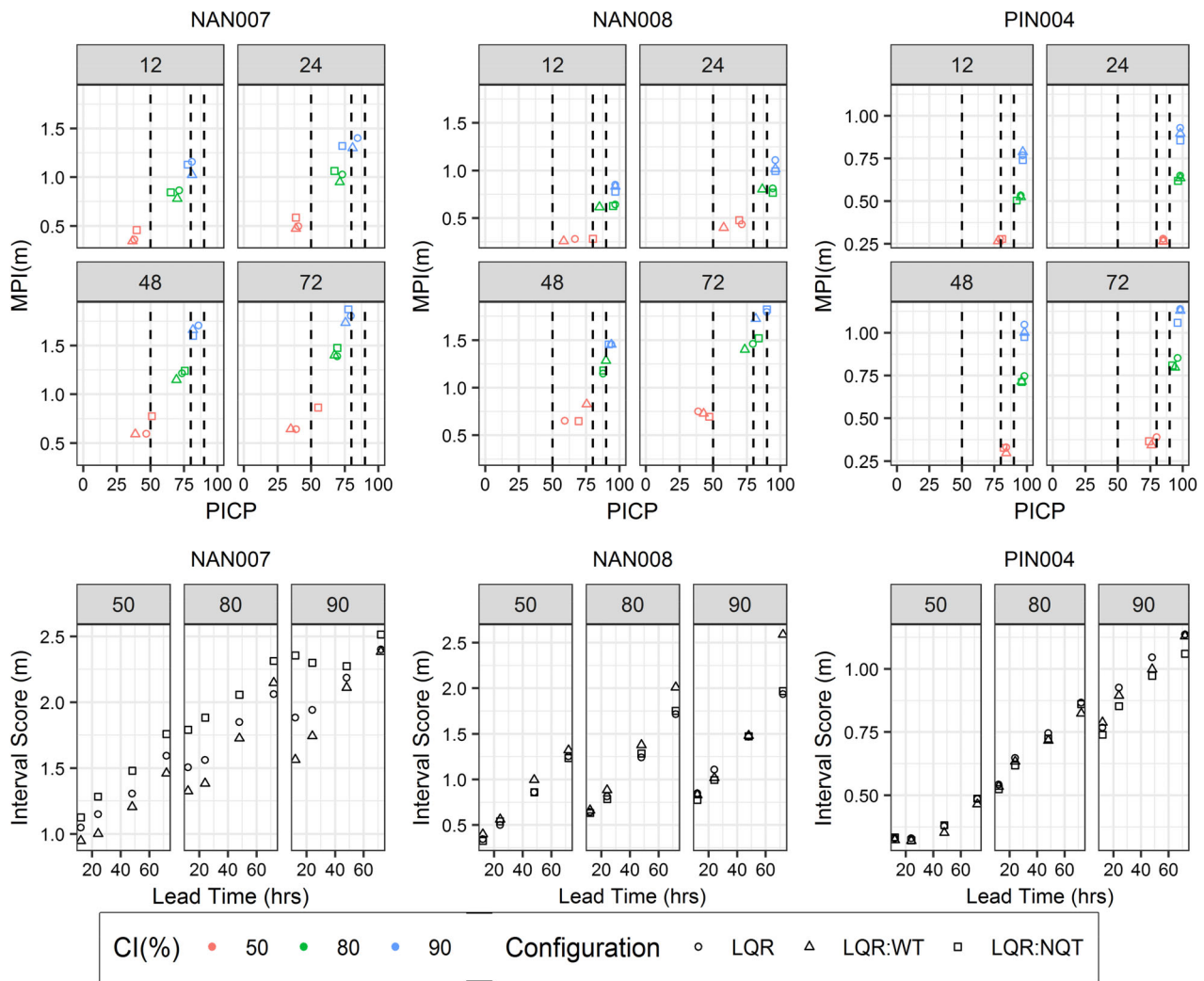
FIGURE 11 PICP-MPI plots (top row) and Interval scores (bottom row) for stations NAN007, NAN008, PIN004 at various lead times and confidence intervals. The colour and shape of scatter corresponds to CI and configuration, respectively. The number in each panel of PICP-MPI plot represents lead time in hours and the Interval score plot represents CI

for LQR:WT for 90% CI but were almost equal for lower CI values. The forecasts at PIN004 were equally unreliable and the interval score solely depend upon MPI for given forecasts. The lowest value for LQR:WT at PIN004 was due to the lowest value of MPI compared to the other configurations. Despite one of the configurations being better in overall Interval score, the forecasts at PIN004 were unreliable, which can be seen in the reliability plot as well as the PICP-MPI plot. The ways in which the forecast can be improved can be assessed using the PICP-MPI plot, which includes both reliability and sharpness and hence, room for improvement can be visually observed.

The overall performance suggests that the predictive uncertainty increases with lead times and it is reflected in the skill scores. Despite increasing uncertainty, reliability plot shows similar performance across lead times

because the reliability does not evaluate the sharpness (width of confidence band). The performance measures which measures sharpness, MPI in this study, is highly affected by the increasing lead times. The metrics which consider both reliability and sharpness shows a deterioration of the score along lead times. This is observed in decreasing CRPSS and increasing MPI with lead times.

5 | CONCLUSIONS

The research demonstrates a comparison of various configurations of quantile regression as a post-processor for the generation of probabilistic hydrological forecasts. Multiple metrics and skill scores were calculated to assess and compare the quality of the generated forecasts. Overall, the performance of the LQR in original and

ACHARYA ET AL.

CIWEM Chartered Institution of Water and Environmental Management Journal of **Flood Risk Management**—WILEY 15 of 16

transformed space was similar to the previous studies. The QR configurations in original space in this study, however, used forecast error as predictand and applied a weighted regression.

The post-processor generates a probabilistic forecast as well as provides a bias correction of the deterministic forecast. Quantile regression in Gaussian space restructures the data and introduces non-linearity in the regression but does not consider extreme events, which lie at the tail of the Gaussian distribution. The application of linear extrapolation for forecasts beyond the range of historical forecasts gives unrealistically high values. However, weighted quantile regression gives importance to less frequent extreme values and can easily extrapolate forecasts beyond the calibration dataset. This reduces the complexity associated with the transformation and assumption of extrapolation at extreme values. Despite showing mixed results for weighted LQR and LQR in Gaussian space, the former approach is therefore recommended.

However, the QR methods have some limitations. The methods develop a direct relation between forecasts and forecast errors. Therefore, it requires long time series of calibration and validation dataset containing multiple extreme events to obtain credible uncertainty estimation models. This limits the model skills for extreme events. Although QR can accommodate heteroscedastic forecast errors, it assumes a linear variation of heteroscedasticity. This also reduces the performance of the QR models analysed in this study.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from Hydro-Informatics Institute. Restrictions apply to the availability of these data, which were used under license for this study.

ORCID
Suwash C. Acharya https://orcid.org/0000-0002-0958-4773

REFERENCES

Arnal, L., Ramos, M.-H., Coughlan de Perez, E., Cloke, H. L., Stephens, E., Wetterhall, F., ... Pappenberger, F. (2016). Willingness-to-pay for a probabilistic flood forecast: A risk-based decision-making game. *Hydrology and Earth System Sciences*, *20*(8), 3109–3128. https://doi.org/10.5194/hess-20-3109-2016

Barbetta, S., Coccia, G., Moramarco, T., Brocca, L., & Todini, E. (2017). The multi temporal/multi-model approach to predictive uncertainty assessment in real-time flood forecasting. *Journal of Hydrology*, *551*, 555–576. https://doi.org/10.1016/j.jhydrol.2017.06.030

Bremnes, J. B. (2004). Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Monthly Weather Review*, *132*(1), 338–347. https://doi.org/10.1175/1520-0493(2004)132<0338:PFOPIT>2.0.CO;2

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, *375*(3–4), 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Dogulu, N., López López, P., Solomatine, D. P., Weerts, A. H., & Shrestha, D. L. (2015). Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrology and Earth System Sciences*, *19*(7), 3181–3201. https://doi.org/10.5194/hess-19-3181-2015

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Han, S., & Coulibaly, P. (2017). Bayesian flood forecasting methods: A review. *Journal of Hydrology*, *551*, 340–351. https://doi.org/10.1016/j.jhydrol.2017.06.004

Koenker, R., & Hallock, K. (2001). Quantile regression: An introduction. *Journal of Economic Perspectives*, *15*(4), 43–56.

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, *249*(1–4), 2–9. https://doi.org/10.1016/S0022-1694(01)00420-6

Krzysztofowicz, R. (2002). Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, *268*(1–4), 16–40. https://doi.org/10.1016/S0022-1694(02)00106-3

Krzysztofowicz, R., & Kelly, K. S. (2000). Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research*, *36*(11), 3265–3277. https://doi.org/10.1029/2000WR900108

López López, P., Verkade, J. S., Weerts, A. H., & Solomatine, D. P. (2014). Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn River: A comparison. *Hydrology and Earth System Sciences*, *18*(9), 3411–3428. https://doi.org/10.5194/hess-18-3411-2014

Montanari, A., & Brath, A. (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, *40*(1), 1–11. https://doi.org/10.1029/2003WR002540

Murphy, A. H. (1993). What is a good forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, *8*(2), 281–293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2

Muthusamy, M., Godiksen, P. N., & Madsen, H. (2016). Comparison of different configurations of quantile regression in estimating

predictive hydrological uncertainty. *Procedia Engineering*, *154*, 513–520. https://doi.org/10.1016/j.proeng.2016.07.546

Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J., & de Roo, A. P. J. (2005). Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European flood forecasting system (EFFS). *Hydrology and Earth System Sciences*, *9*(4), 381–393. https://doi.org/10.5194/hess-9-381-2005

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., ... Salamon, P. (2015). How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology*, *522*, 697–713. https://doi.org/10.1016/j.jhydrol.2015.01.024

Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, *133*(5), 1155–1174. https://doi.org/10.1175/MWR2906.1

Ramos, M. H., van Andel, S. J., & Pappenberger, F. (2013). Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences*, *17*(6), 2219–2232. https://doi.org/10.5194/hess-17-2219-2013

Roscoe, K. L., Weerts, A. H., & Schroevers, M. (2012). Estimation of the uncertainty in water level forecasts at ungauged river locations using quantile regression. *International Journal of River Basin Management*, *10*(4), 383–394. https://doi.org/10.1080/15715124.2012.740483

Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, *19*(2), 225–235. https://doi.org/10.1016/j.neunet.2006.01.012

Sisomphon, P., Boonya-aroonnet, S., & Chonwattana, S. (2013) Towards the development of a decision support system for flood management in Chao Phraya River Basin, Thailand in *International Conference on Flood Resilience*, *Experiences in Asia and Europe*, *5–7 September* 2013; *Exeter*, *UK*.

Solomatine, D. P., & Shrestha, D. L. (2009). A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, *45*(12), 1–16. https://doi.org/10.1029/2008WR006839

Thordarson, F. Ö., Breinholt, A., Møller, J. K., Mikkelsen, P. S., Grum, M., & Madsen, H. (2012). Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion. *Stochastic Environmental Research and Risk Assessment*, *26*(8), 1151–1162. https://doi.org/10.1007/s00477-012-0563-3

Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management*, *6*(2), 123–137. https://doi.org/10.1080/15715124.2008.9635342

Todini, E. (2009). Predictive uncertainty assessment in real time flood forecasting. In *Uncertainties in environmental Modling and consequences of Plicy making* (pp. 205–228). Dordrecht: Springer. https://doi.org/10.1007/978-90-481-2636-1_9

Todini, E. (2017). Flood forecasting and decision making in the new millennium. Where are we? *Water Resources Management*, *31*(10), 3111–3129. https://doi.org/10.1007/s11269-017-1693-7

Van Steenbergen, N., Ronsyn, J., & Willems, P. (2012). A nonparametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling & Software*, *33*, 92–105. https://doi.org/10.1016/j.envsoft.2012.01.013

Verkade, J. S., & Werner, M. G. F. (2011). Estimating the benefits of single value and probability forecasting for flood warning. *Hydrology and Earth System Sciences*, *15*(12), 3751–3765. https://doi.org/10.5194/hess-15-3751-2011

Weerts, A. H., Winsemius, H. C., & Verkade, J. S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences*, *15*(1), 255–265. https://doi.org/10.5194/hess-15-255-2011

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences International Geophysics Series* (Vol. 100). New York, NY: Academic Press.

Author/s:
Acharya, SC; Babel, MS; Madsen, H; Sisomphon, P; Shrestha, S

Title:
Comparison of different quantile regression methods to estimate predictive hydrological uncertainty in the Upper Chao Phraya River Basin, Thailand

Date:
2020-03

Citation:
Acharya, S. C., Babel, M. S., Madsen, H., Sisomphon, P. & Shrestha, S. (2020). Comparison of different quantile regression methods to estimate predictive hydrological uncertainty in the Upper Chao Phraya River Basin, Thailand. Journal of Flood Risk Management, 13 (1), https://doi.org/10.1111/jfr3.12585.

Persistent Link:
http://hdl.handle.net/11343/252457

File Description:
Published version
License:
CC BY