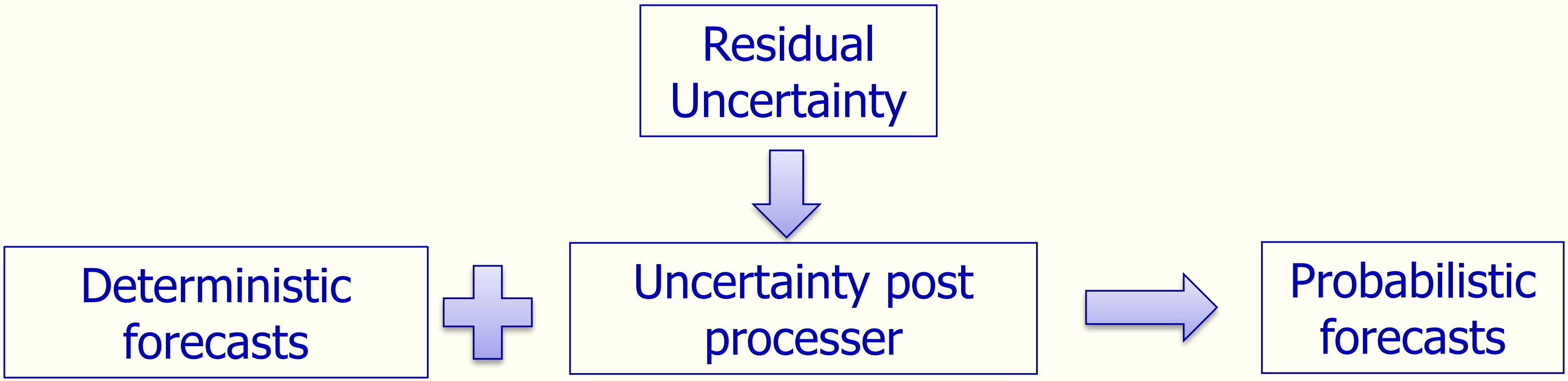
Different configurations of quantile regression in estimating predictive hydrological uncertainty

Manoranjan Muthusamy [a,b], Peter Nygaard Godiksen [b], Henrik Madsen [b]

[a] University of Sheffield, UK
[b] DHI, Denmark

Background



e.g.

- Water level
- Discharge

e.g.

- Hydrological Uncertainty Processor (HUP) *(Krzysztofowicz, R. et al , 2000)*
- Bayesian Model Averaging (BMA) *(Raftery et al., 2010)*
- UNcertainty Estimation based on local Errors and Clustering (UNEEC) (*Solomatine and Shrestha, 2009),*
- Quantile Regression (QR) *(Weerts et al. 2011)*

Method of estimating conditional function of variable of interest (Forecast error in this case) for all quantiles of a probability distribution

Progress so far

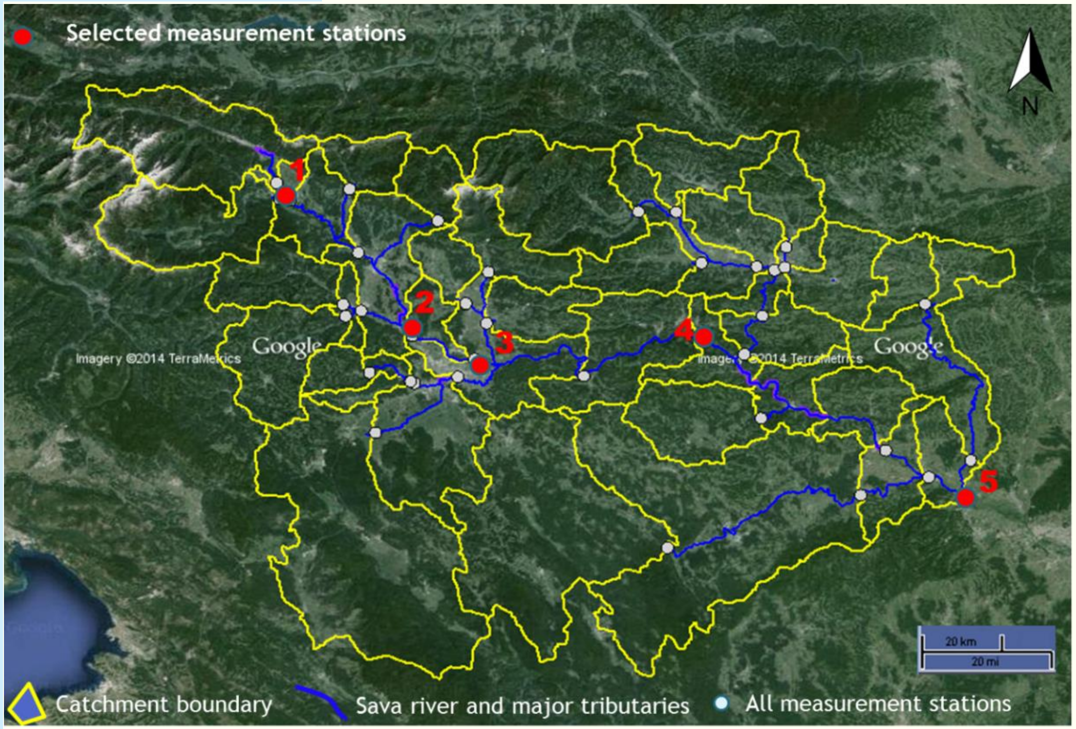
Recent studies have attended to move from linear to non linear quantile regression using

1. Quantile regression in Gaussian domain *(Weerts et al. 2011)*
2. Piece wise linear quantile regression *(López López et al. 2014)*

Objectives of this study

- Comparison of quantile regression in original domain (QR-ORI) vs quantile regression in Gaussian domain (QR-NQT) in the context of flood forecasting.
- Introducing weights during linear quantile regression (QR-WT) to emphasize more on a high flow and compare the performance against QR-ORI and QR-NQT

Real time flood forecasting system Sava River, Slovenia



Common features	
Hydrological and hydrodynamic modelling	MIKE 11
Length (Sava River)	188km
Branches (up to 2nd order tributaries)	23
Basins	40
Discharge measurement stations	22

Statistical properties of observed discharge data (m^3/s)

ID	Chainage (m)	Period	Mean	Max	Min
1	850	22-Dec, 2009 to 23-Sep, 2013	46.9	802	5.34
2	41455		89.8	1292	24.8
3	54405		92.2	1349	23.1
4	108424		175	2152	38.3
5	173000		281	3837	51.9

Data (From – To)	Application
Nov,2011 - Sep,2013	Training
Dec,2009 – Oct, 2011	Validation

Quantile regression

- Quantile regression: Method of estimating conditional function of variable of interest (Forecast error in this case) for all quantiles of a probability distribution

- Ordinary least square (OLS) - Finds the sample mean by minimizing the sum of squared differences
  Quantile regression (QR) - Finds the particular quantile by minimizing the sum of asymmetrically weighted absolute residuals $(u = y_i - \xi,$ *where* $y - variable\ of\ interest,\ \xi - Quintile\ regression\ function)$

$$min \sum_{i=1}^{n} \rho_\tau u$$

Where $,\rho_\tau = \begin{Bmatrix} (\tau - 1) \cdot u, & u < 0 \\ \tau \cdot u & u \geq 0 \end{Bmatrix}$

- Conditional quantile regression - describes quantiles depending on covariate (x_i)

$$min \sum_{i=1}^{n} \rho_\tau(y_i - \xi(x_i, \beta))$$

Where, $\beta - vector\ of\ the\ parameters\ of\ the\ regression$

Quantile regression

- Major steps
  - Conditional quantile regression is derived using,

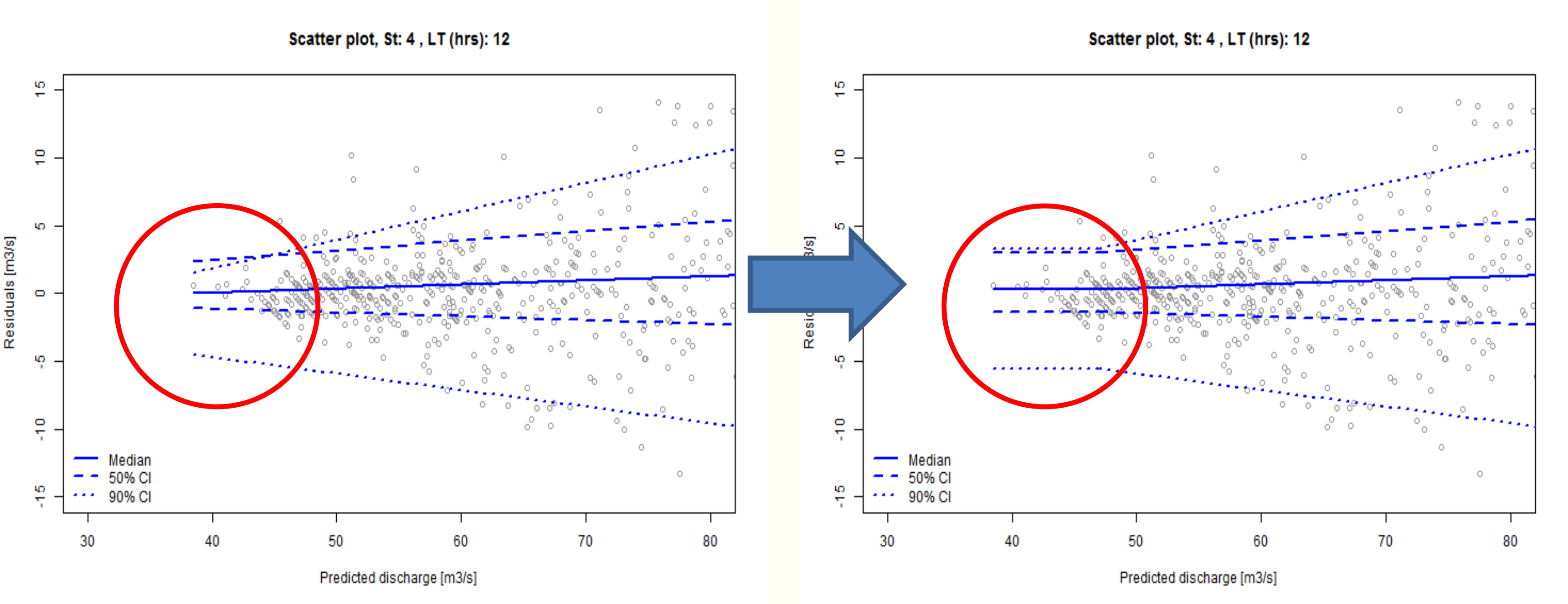
    $min \sum_{i=1}^{n} \rho_{\tau}(e_i - (a_{\tau}\overline{s_i} + b_{\tau}))$

    Where, covariate, $\overline{s_i}$= forecast discharge (m^3/s)
    dependent variable, e_i = deterministic error (m^3/s)
    a_{τ} , b_{τ} = parameters of the linear regression
  - From the conditional quantiles, probability distribution of error conditioned on the forecast discharge is estimated for each lead time using training data set
  - This model is applied as a post processer of deterministic forecasts in validation period for the lead time of interest

- Three different approaches are tested
  1. Quantile regression in original domain (QR-ORI)
  2. Quantile regression in Gaussian domain (QR-NQT) *(Weerts et al. 2011)*
  3. Weighted quantile regression in original domain (QR-WT)

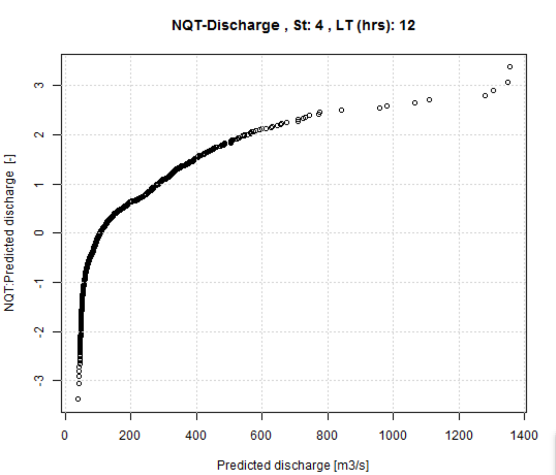
Quantile regression in original domain (QR-ORI)

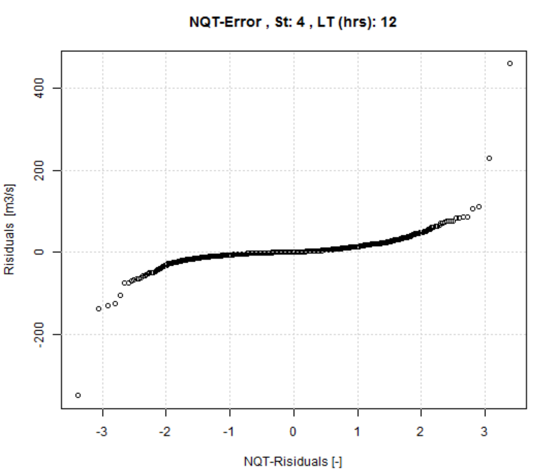
- Crossing of quantiles solved by defining a constant error model below this level
  - simple yet feasible solution
  - effects only a small portion of low flow



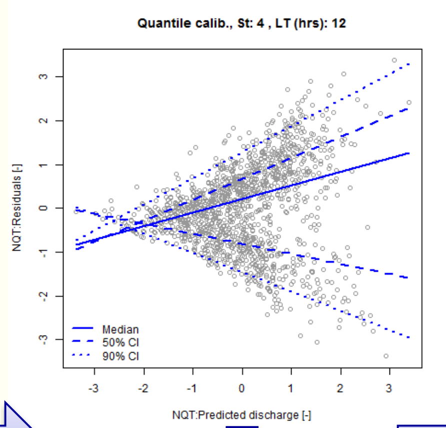
Quantile regression in Gaussian domain (QR-NQT)

Normal Quantile Transformation (NQT)

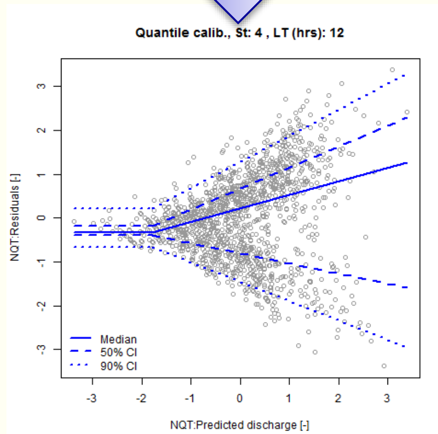




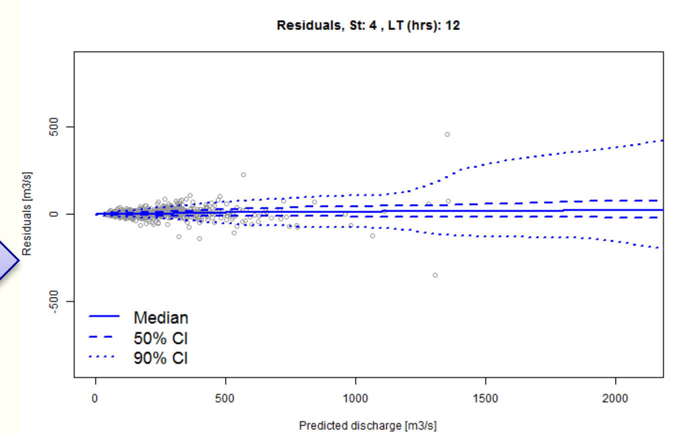
Quantile regression in Gaussian domain



Non- crossing



Back transformation to original domain



Weighted quantile regression (QR-WT)

- Higher weights are given to higher discharge to take advantage of better calibration of hydrological modelling at high flow

- Conventional quantile regression : Regression lines fits to minimize the sum of the absolute residuals
  Weighted quantile regression : Regression lines fits to minimize the sum of the weights multiplied into the absolute residuals

- Weight of a random forecast discharge ($\bar{s}_i$) , $w_i = \frac{r_i}{N}$

  r_i - Rank of $\bar{s}_i$,
  N –Total number of samples

Verification measures

- Brier Score (BS) – measures the mean squared error of a probabilistic forecast

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

n -Number of pairs of forecasts and observations,
f_i -Predicted probability of forecast i
o_i -1 or 0 (event occurred or not)

- Performance of QR-ORI, QR-NQT and QR-WT are compared using prediction interval coverage probability (PICP) and mean prediction interval (MPI) *(Shrestha and Solomatine 2006)*

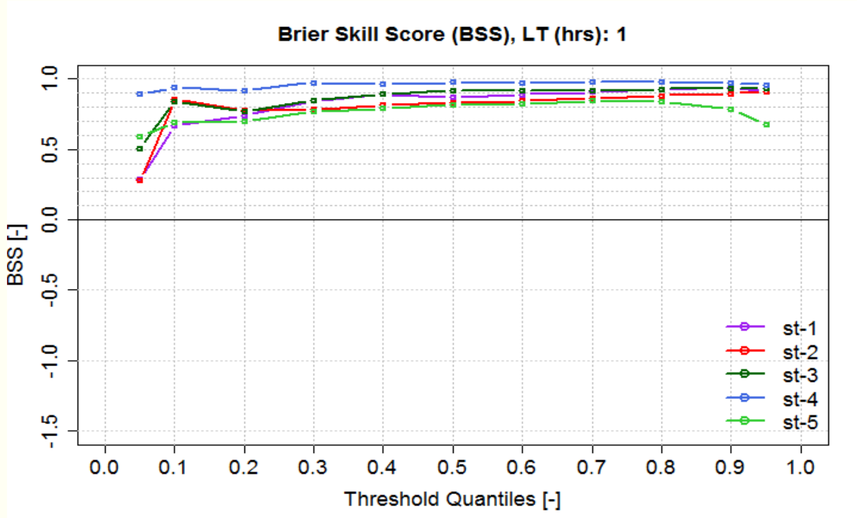
$$PICP = \frac{1}{n}\sum_{i=1}^{n} R * 100\% \qquad where\ R \begin{cases} 1, & PL_i^u \leq O_i \leq PL_i^l \\ 0, & otherwise \end{cases}$$

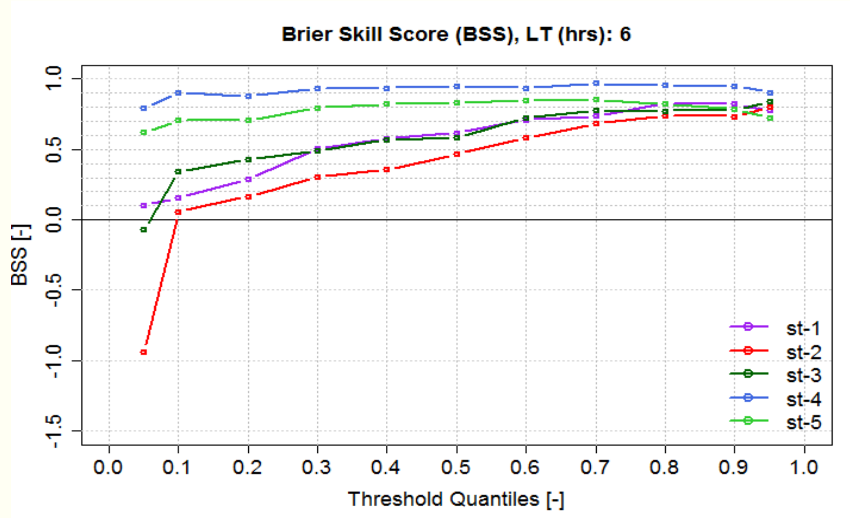
$$MPI = \frac{1}{n}\sum_{i=1}^{n}(PL_i^u - PL_i^l)$$

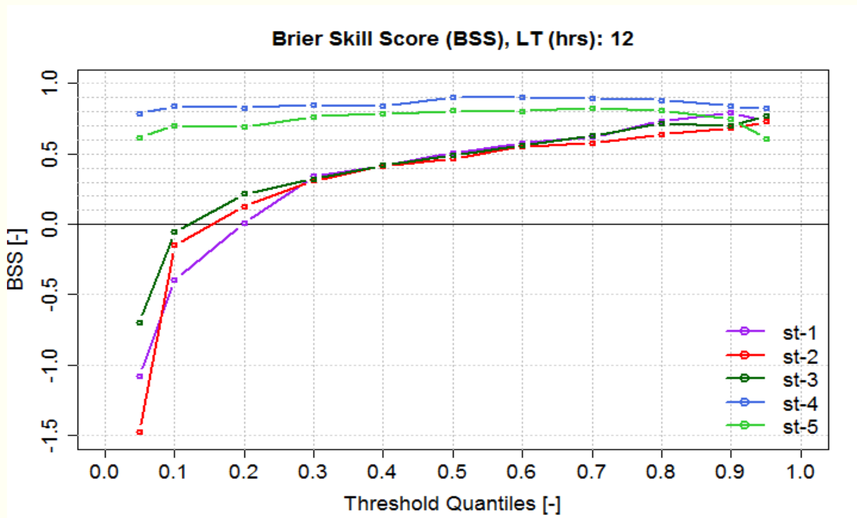
PL_i^u, PL_i^l - upper and lower boundary of the considered confidence interval at time, *i*
O_i - observed discharge at time, *i*

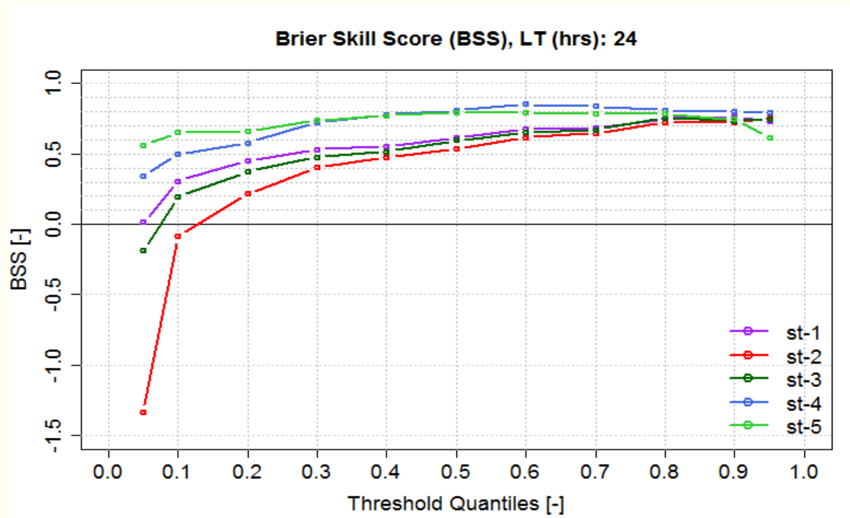
PICP - Measures reliability (%)
The closer to considered prediction interval (5%,10%,….,90%, 95%) the better
MPI - Measures resolution (m^3/s)
The smaller the better

Brier Skill Score – QR (ORI)

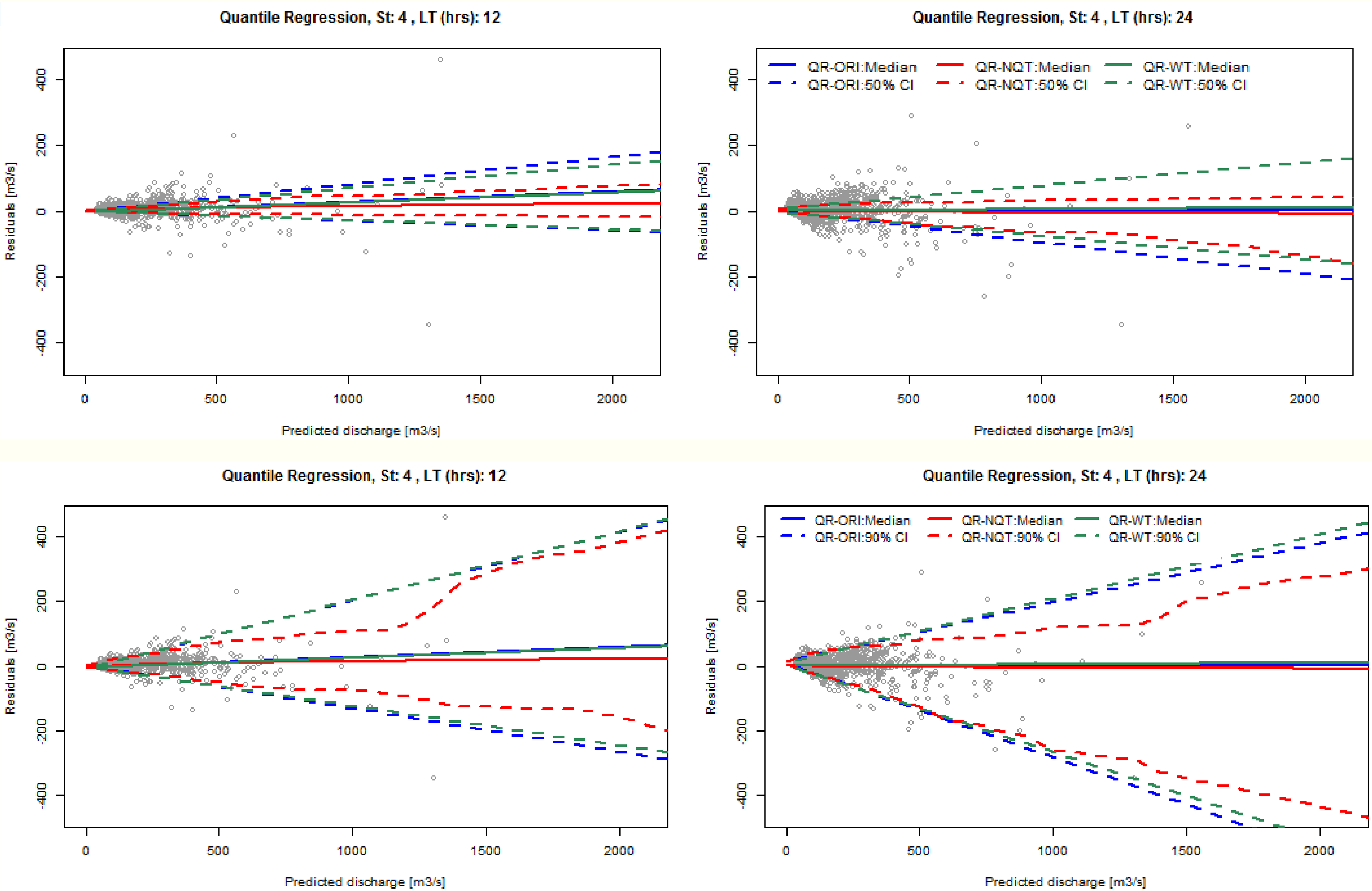






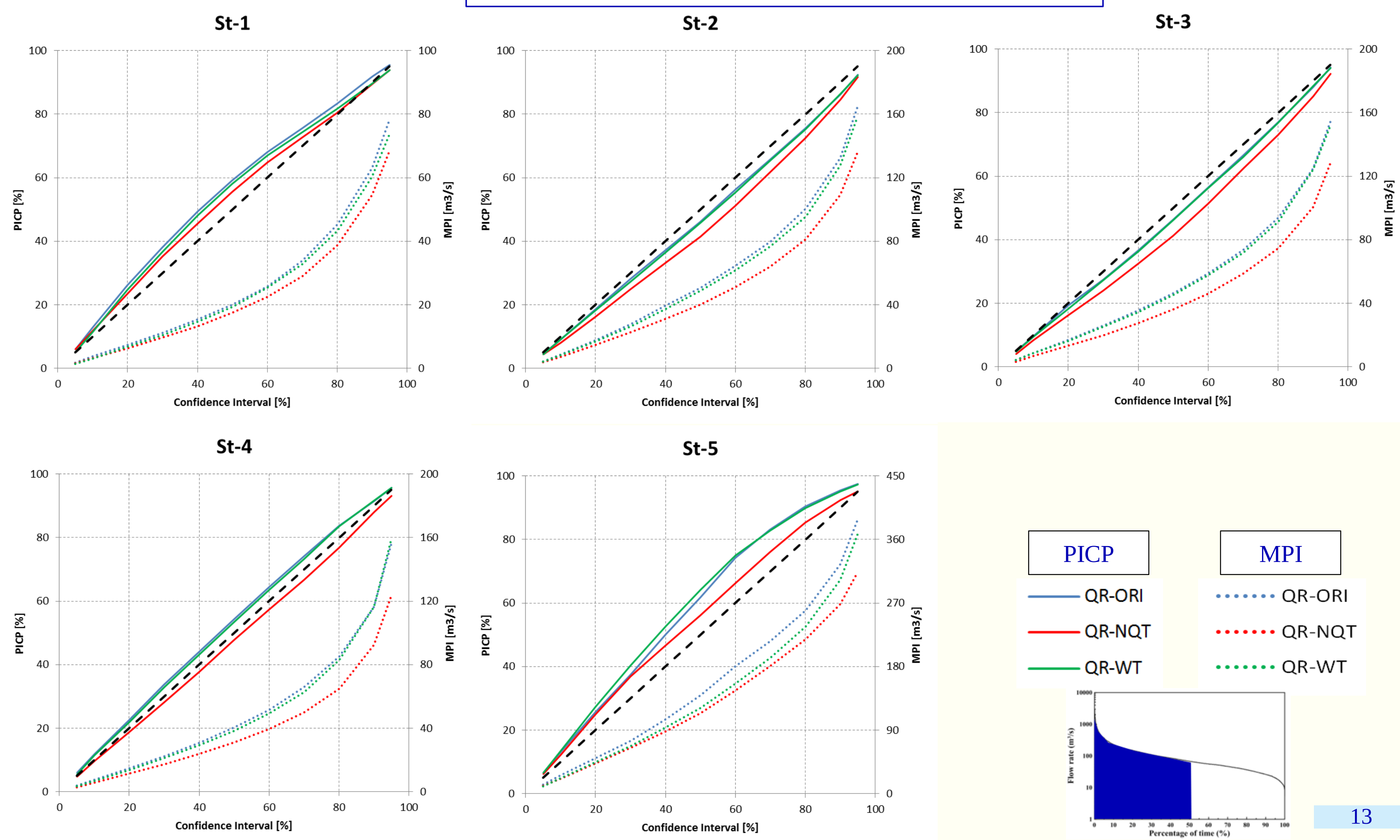


Comparison of QR-ORI,QR-NQT and QR-WT



Comparison of QR-ORI,QR-NQT and QR-WT

PICP-MPI plots – Highest 25% data



- QR-NQT outperforms QR-ORI in terms of forecast resolution for all stations for the highest 25% of the forecast discharge . But in terms of reliability, improvements are largely depending on the size of the training and validation data set and the range of discharges in both data set

  -A comparison between QR-NQT and QR-ORI with large training and validation data set consist of wider distribution of values is recommended

- QR-WT shows slightly better performance than QR-ORI specially in terms of resolution for highest 25% of the forecast discharge which is the flow regime of interest in flood forecasting

  -Weighted quantile regression can also be applied in Gaussian domain. By doing this, while more emphasize is given to higher flow, non-linear quantile regression relationship can be derived in original domain

- Overall the probabilistic forecasts derived using quantile regression method show good skills considering Brier skill score (BSS)

  -Crossing problem of quantiles is solved by defining a constant error model. But a detailed study on other possible solutions is recommended ((López López et al. 2014)

  -Comparison of the performance with another uncertainty predictor (e.g. UNcertainty Estimation based on local Errors and Clustering (UNEEC)) is recommended (Dogulu et al. 2015)

References

[1] R. Krzysztofowicz, K.S. Kelly, Hydrologic uncertainty processor for probabilistic river stage forecasting, Water Resour. Res. 36 (2000) 3265–3277.

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

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

[4] P. Lopez Lopez, J.S. Verkade, A.H. Weerts, D.P. Solomatine, Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn River: A comparison, Hydrol. Earth Syst. Sci. 18 (2014) 3411–3428. doi:10.5194/hess-18-3411-2014.

[6] D.L. Shrestha, D.P. Solomatine, Machine learning approaches for estimation of prediction interval for the model output., Neural Netw. 19 (2006) 225–35. doi:10.1016/j.neunet.2006.01.012.

[7] Dogulu, N., López López, P., Solomatine, D. P., Weerts, A. H., and 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., 19, 3181-3201, doi:10.5194/hess-19-3181-2015, 2015.

Acknowledgement

(a) European Union's Erasmus Mundus Joint Master Degree (EMJMD) programme and

(b) Marie Curie ITN (Quantifying Uncertainty in Integrated Catchment Studies) which is a part of European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000

Thank you!