Hydrol. Earth Syst. Sci. Discuss., 11, C4049–C4054, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C4049/2014/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Estimation of predictive hydrologic uncertainty using quantile regression and UNEEC methods and their comparison on contrasting catchments" *by* N. Dogulu et al.

Anonymous Referee #1

Received and published: 27 September 2014

1. General Summary

This paper is generally well written and addresses an operationally important problem such as the application of uncertainty processors in flood forecasting. In this context, the authors present a comparison between two existing methods for uncertainty assessment, Quantile Regression (QR) and UNEEC. Even though the main topic of the paper could be an informative contribution to the hydrological literature, overall the structure of the paper is quite confused, especially regarding the experimental setup and the analysis of the results. In particular, the latter is not enough in-depth and often contradictory. The comparison of the two methods is not carried on rigorously

and the evaluation indexes used to compare the results of the two methodologies are often misinterpreted. In my opinion, these gaps preclude the paper to be a novel contribute to the hydrological literature and helpful in the choice between the two methods in operational applications.

2. Main Comments

1) The comparison of the two methods is not enough rigorous. In fact, since I do not see any limitations regarding the number and typology of predictors to be used in both methods, the comparison should have been done using the same predictors for both estimators; otherwise, the effect of a different information level used to force the estimators becomes more significant than the differences in the methods. QR is always used with the only model prediction as a predictor, while UNEEC includes observed values at previous time steps and state variables provided by the hydrological model (such as ground water lever and soil moisture deficit). The authors, in Chapter 5 (page 10208, lines 14-15), confirm this saying that introducing more predictors in the QR methodology could possibly increase the performance of QR, assuming that the conclusions of the comparisons are affected by the choice of different predictands. Moreover, it is not clear to me why the hydrological model prediction has not been used as a predictor in the UNEEC setup on both rivers, the authors did not explain this choice. In my opinion, the authors should have carried out the comparison using the same predictors or at least giving a convincing explanation for the choice of different predictors, otherwise the result of the comparison are obviously biased towards the estimator forced with the better information. The paper in its current form shows a misunderstanding of uncertainty assessment capability of the methodology and informative level of the predictors.

2) In Section 2.2 (page 10191, lines 24-27), the authors claim that "none of the presented measures allow for accurate comparison between different methods of uncertainty prediction and should be therefore seen only as indirect indicators of method's performance". In my opinion, this statement is incorrect. In fact, the PCIP is often

enough to evaluate the correctness and performances of an uncertainty estimator because it verifies whether the estimated uncertainty distribution is correct (i.e., includes the right amount of observed data) or not. Once this is proved, the other indexes (MPI and ARIL) can be used to understand how wide the uncertainty is and if it may be reduced using different predictors. The authors also point out correctly that ARIL may be affected by misleading values when the streamflow is 0 or very small. In order to evaluate how much ARIL is affected by these values the reader should have a better idea of the streamflow distribution of the case studies, but the authors only provide the mean flow making the interpretation of ARIL very difficult. Moreover, when the streamflow is close to 0 the uncertainty is usually pretty small (compared to the average value of the uncertainty band width), so it would have been helpful to screen out these values, which do not have a significant impact in the analysis, when computing the index. The wrong interpretation of the indexes led to some arguable conclusions:

a) Page 10202, lines 14-15. The authors say that "QR produces unnecessarily wider uncertainty bounds for medium peaks in validation". The fact that the uncertainty band is unnecessarily wider should be proved showing the PCIP for the cluster including medium events. The authors do not show these indexes for the validation period so the reader can only rely on Table 3, which shows the indexes for the training period. According to this table, QR has a PCIP often lower than 90% or very close and only for cluster #3 it is significantly higher, but also UNEEC for that cluster gives a high value of PCIP. From Table 2, QR shows lower PCIP values during validation that those computed during training, so I suppose (maybe wrongly) that the same happen for most of the clusters. This would lead to think that the PCIP in validation is always lower than 90% for all the clusters and this would be in contrast with what the authors claim about the unnecessarily wide bounds.

b) Page 10202, lines 20-22. The authors write that from Table 3 it is possible to verify a contradictory relation between PCIP and MPI, since the latter is higher when the former is closer to 90%. In Table 3 I do not see any contradictory relation, because for every

cluster MPI is higher when PCIP is higher and this just implies that a wider uncertainty band includes a higher number of observed values.

c) Page 10203, lines 3-4. The authors say that for the cluster of high flow the NUE index must be considered to correctly compare QR and UNEEC and they conclude that UNEEC performs better because it yelds a higher NUE value. This statement is misleading because it seems that the authors compute NUE for better analyzing the cluster #4, but then they point out general conclusion for every cluster based on that index. Moreover, the author themselves claimed in page 10191, lines 21-23 that a higher NUE does not imply a better performance and I completely agree with this statement. However, they are now contradicting their words using a higher NUE as simple evidence of the UNEEC better performance, without considering, for example, the fact that for clusters 4 and 5 the PCIP given by UNEEC is very far from 90% if compared to that given by QR.

d) Page 10203, lines 9-11. The low values of MPI in Yeaton catchment do not surprise me mainly because the mean flow (as reported in Table 1) is much lower than that of the other catchments and, in smaller part, also for the fact that the hydrological model is more accurate. I would rather use ARIL for this analysis, because it accounts for the flow magnitude. Actually, if one considers ARIL the situation is different, Yeaton has the worst value for 24-hr lag time and it has a value higher than that of Llanerfyl for all the others lag times.

e) Page 10203, lines 25-26. The author claim that QR does slightly better than UNEEC in Yeaton. I would rather say that UNEEC performs very poorly on this catchment considering the extremely low values of PCIP when the 50% uncertainty band is considered.

f) Page 10204, lines 6-7. I agree with the authors regarding the 90% uncertainty band, but I disagree for the 50% band since UNEEC gives very low PCIP for the lower time lags.

g) Page 10204, lines 8-9. It does not seem so clear to me, especially for the 50% band.

h) Page 10205, lines 13-19. From Figure 14 it is almost impossible to see that UNEEC prediction intervals are wider. However, the explanation of the reason why UNEEC provides wider intervals is not clear to me.

i) Page 10206, lines 24-26. This sentence is not clear at all. Slightly better values of PCIP compared to cluster 2 or to QR? Why can they be attribute to lower MPI?

j) Page 10208, lines 1-5. I do not agree with the authors when they claim that there is no basis for comparison of different uncertainty estimators. In my opinion, PCIP is the basis, if it is far from the expected value the estimator is not reliable and useless in real application. In case both estimators gives good PCIP values, then the PCIP for different conditions must be analyzed (e.g. for different clusters) to check if the correctness of the estimator is preserved for different situations. If still the methods give similar results, then the MPI and ARIL can be used to identify the better methodology.

3) The evaluation of the methods performance should mainly focus on the PCIP and Q-Q plots (Laio and Tamea, 2007) of both calibration and validation data. The authors often show only analysis of training data (Tables 3 and 4; Figures 5, 6, 7, 8, 9 and 10) and sometimes only of validation data (Figure 12 and 13). Only in Table 2 training and validation periods are showed together. A comparison of both periods is necessary to evaluate the ability of an estimator to evaluate the uncertainty of new/unknown data, which is fundamental in real time applications. Moreover, in Figures 7b and 10b only the cluster with the distribution closest to normal is showed; I would rather show the cluster with the distribution furthest to normal (or at least both) to better understand the origin of the error in the uncertainty assessment.

4) Section 3.2 is very confused. It is not always clear if the authors are referring to the case study of Brue or to the Upper Severn catchments. The description of the experimental setup is mixed up with few analyses of the hydrological model performances, which are not necessary for the purpose of the section (Pages 10198-10199, lines 25-

3). The description of the choice of the predictors for the Uppern Severn catchments is not very clear and linear as it should be. Some sentences (e.g. “low soil moisture is more likely attributed to higher rainfall rates”) show a lack of effort in making this section easily understandable.

3. Minor Comments

1) Page 10187, lines 17-20. The authors do not explain why they used the variant called QR1.

2) Page 10196, lines 6-7. The sentence is not clear to me, maybe “are obtained” should be replaced with “is obtained”

3) Page 10197, lines 18-19. The sentence “low soil moisture is more likely attributed to higher rainfall rates” does not make much sense.

4) Page 10198, line 4. “et” should be replaced by “e” or “et-i”

4. References

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

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 10179, 2014.