Response to review comments of Anonymous Referee #2
on the manuscript "Estimation of predictive hydrologic uncertainty using quantile regression and UNEEC methods and their comparison on contrasting catchments" ***by Dogulu et al. 2014***

We would like to thankfully acknowledge Referee #2 for her/his thorough review and valuable comments. We believe that addressing these comments will help improving the quality of the original manuscript. Below we give a point-by-point answer to the individual comments.

RC: *Even though the topic is definitively of interest, the novel contribution of the proposed manuscript for the advancement of the knowledge and for practical purposes is rather limited.*

AC: We realize that that our manuscript, focusing at the comparison of only two methods for predicting uncertainty, provides only a limited contribution for advancement of knowledge. Still we hope that since the referee thinks that "the topic is definitely of interest" this paper addressing the methods of residual uncertainty prediction (rather than much more "popular" parametric uncertainty) may be worth reading by the audience of HESS (given all justified comments are addressed by the authors of course). We felt the comparison of UNECC to other methods in an earlier paper have not gone far enough. After the first round of reviews we of course realize that the analysis has to be much better structured and certain formulations and logical constructs require amendment and/or refining.

RC: *Indeed, the author state that "The motivation here is to identify possible advantages and disadvantages of using QR and UNEEC methods based on their comparative performance, especially during flooding conditions (i.e. for the data cluster associated with high flow/water level conditions)", but the analysis performed are rather confusing and lead to some questionable conclusions.*

AC: We will critically review all conclusions made (giving special attention to the "questionable" ones) and revise them to improve clarity and logic.

RC: *In short, the lack of a significant novel contribution and the questionable analysis and conclusion provided affect the quality of the proposed manuscript making it not a valuable contribution for the advancement of knowledge.*

AC: Please see the previous answers.

RC: *Sect. 4.1 statistical error analysis: it seems that omoscedasticity/ heteroscedasticity of the error is an important aspect affecting the performances of the models; with respect to this, the authors could considered the comments provided in Coccia and Todini (2011) concerning the QR.*

AC: We thank the referee for the suggestion. Based on the reference given by the referee we will extend our discussion both on the QR (Sect. 2.1.1) and the model error heteroscedasticity (Sect. 4.1) where possible.

RC: *p.10202, line 20 and Table 3: why only training results are presented? What about validation results?*

AC: This issue is also raised by the Anonymous Referee #1. We did not present the results from the validation mainly due to the methodological basis that defines the framework of the UNEEC method. More details can be followed from the author response to the relevant comment by the Referee #1.

RC: *p.10202, lines 20-22: why is the relationship between PICP and MPI contradictory?*

AC: The term "contradictory" is a bit strong. The reason we mention this is the following: A higher PICP (closer to confidence level specified) and lower MPI (and ARIL) value means that a particular method estimates predictive uncertainty more accurately. However, intrinsic to their definitions, there is often a compromise between the validation measures PICP and MPI (and ARIL) when they are used together in evaluating performances of different methods: better PICP values in terms of closeness to the specified confidence level can be obtained with larger MPI (and ARIL) values. We will include the relevant explanations discussion in "Sect. 2.2 Validation methods".

RC: *p.10203, lines 2-4: I disagree with the final conclusion provided by the author about the performances of the two compared method with respect to the Brue catchment. In fact, they state that in general UNEEC shows a better performance since it yields a higher NUE value, but it is worth noting that an uncertainty estimation method should first of all provide correct PICP values (and being the PICP correct, lower MPI or ARIL values are worth wishing for); indeed, according to Table 3, UNEEC generally provides less correct PICP values than QR, particularly for cluster 4: it is worth noting that this cluster includes high flows/high rainfalls, that, according to the authors, are those of major concern.*

AC: The authors thank the referee for pointing this out. Indeed the PICP is the primary indicator. NUE (Nasseri and Zahraie, 2011) is an update of PICP (ARIL is added in denominator) and indeed has to be seen as an additional indicator. So PICP-wise, QR is better, and NUE-wise, however, UNEEC is better than QR. In our view it is worth looking at several indicators. In the revised manuscript this point will be properly addressed.

RC: *p. 10205, lines 13-24: it is almost impossible observing in figure 14 the considerations and comments provided in the manuscript; please modify the figure making it more clear.*

AC: We will improve the quality of this figure for the revised manuscript.

RC: *p. 10206 and Table 4: as for Table 3, why only training results are presented? What about validation results?*

AC: Please see above the response to referee comment on Table 3.

RC: *p. 10207-10208 Sect. Conclusions: a clear and robust conclusion addressing the problem presented in the introduction (see my general comment) is missing.*

AC: Agreed. Both the Introduction and the Conclusion parts (as well as the Abstract) will be revised significantly in light of the relevant comments by the referees and the author responses.

RC: *p. 10208, lines 8-15: this aspect could be of interest, but the hypothesis is supported just by one case study. It is worth noting that in the study by López López et al. (2014) (featuring several authors in common with this one and that, according to the authors themselves, is accompanying it – see page 10207 line 7) many other Severn catchments, most of them characterized by low lag times, were considered (see Table 1 of López López et al. (2014)): these catchments could be considered to support and make more robust the hypothesis provided.*

AC: This study considers three catchments which are distinctively representative of the other Upper Severn catchment with respect to their basin lag time as well their mean flow. We do agree with the Referee #2 that several more Upper Severn catchments characterized by low lag times (i.e. Llanymynech, Vyrnwy Weir, Bryntail, and Rhos Y Pentrof) could have been considered to make the hypothesis we suggest for the rapid response catchments more robust. Nevertheless, by the time we put this hypothesis forward there were only limited time available left for the Erasmus Mundus Flood*Risk* Master students (Ms. Nilay Dogulu and Ms. Patricia López López) to complete their MSc research (which lasts only six months) and continue with their career elsewhere. Consequently we could not study any further rapid response catchment in the region. Still we do think that the results from these three case studies have enabled us to reach conclusions that can be useful for the

operational hydrology community. If resources and time is given to continue this research we will certainly extend the number of case studies to strengthen out conclusions.

RC: *p.10186: Q should be theta*
AC: The referee is right. This will be changed in the revised manuscript.

RC: *p. 10191, lines 18-20: "Nasseri and Zahraie (2011) recommend that methods with the higher NUE should be preferred over those with the lower NUE: : :". I did not find such a recommendation in Nasseri and Zahraie (2011)*
AC: Agreed – the authors' mistake in citation. The correct citation is "Nasseri and Zahraie (2013)" and will be corrected in the revised manuscript. We also would like to note that the sentence in line 15 on the same page will be updated as follows: "A possibility to combine PICP and ARIL is to use the NUE indicator proposed by Nasseri and Zahraie (2011):".