1- The paper lacks a discussion of the consequences of the data processing responsible for the different results. E.g. a classification with data recognizing catchment size has to be different of a classification with the same data and considerably less influence of the catchment size (as you get with normalization by average runoff). The meaning of a specification for the data, e.g. elimination of outliers, noise, unexpected runoff behaviour, range of dispersion or loss of information, is an important task for interpreting the results. Understanding the consequences of data specification is essential! A general recommendation for one specification, regardless the issue of the classification, cannot be made for a not completely understood Specification (PredF). Therefore, please consider the meaning of the specifications in the discussion and conclusion.

In this paper we analysed two types of specification that can affect classification characteristics: 1) The data processing (raw versus normalized flow series) and 2) the classification procedure (ClasF vs PredF). In addition, we analysed how these two specifications affect four classification characteristics: 1) classification performance, 2) hydrological interpretation of classification, 3) ability to deal with underrepresented parts of the hydrological space and 4) The spatial correspondence.

According to Referee's comment we have emphasized in the introduction and especially in the discussion and conclusions how each type of specification can affect the different classification characteristics. We included and changed many sentences in the new version of the manuscript where we addressed specifically the effect of those specifications to clarify the effect of each specification.

Within this comment, Referee said that "a general recommendation cannot be made for a not completely understood specification (PredF). After reanalysed several issues according to the comments of Referee #1 and #2 results showed that PredF presented higher classification performance and a greater ability to estimates the hydrological character of the underrepresented parts of the hydrological space. In addition, after comparing our classification with others that covered part of the study area (Bejarano et al., 2010; Solans and Poff, 2013), we could concluded that PredF generated classes that are more similar to the actual distribution of river types in the study zone than ClasF. Then, we believe that according to our analysis and results we now understand how the PredF strategy works and why it outperformed ClasF. Therefore, we change the sentence in the conclusion.

2- The PredF strategy should be explained more in depth to understand what is done with the data and what the consequences for the data are. It seems that the PredF strategy leads to a loss of information with respect to the variability of data. The resulting data may not cover the whole real data space. The advantage of PredF is the possibility to construct data for underrepresented conditions to obtain classes of equal

size. However, to find classes of equal size is not a priority objective of clustering and classification. Please comment on this.

In regard to the loss of information mentioned by the referee, we have modified the manuscript to clarify what we have done with the Synthetic River Network (SRN). The misunderstanding of the procedure was probably due to the position of the paragraph in the original version. We have put it before explaining any of the two strategies to clarify what has been done with the data. The reduction of the SRN from 667406 to 178296 has been done in both ClasF and PredF. This has been done because there are certain types of rivers that are not represented in our initial data set due to the absence of gauges in these rivers. Therefore we limit the potential of our predictions to rivers represented in our initial data set which was defined according to the range of the different predictor variables. However, as stated, it has been done for ClasF and PredF, so results extracted from each strategy are comparable.

Referee is right in saying that finding classes of equal size is not a priority objective of classifications. In figure 7 we included a line representing the “theoretical” most even distributed class (i.e. if all the classes of a classification have equal size). This line was only included as a reference benchmark to see the frequency of the classes that incorporated the distinctive gauges. However, it does not indicate that all the classes should have this frequency. We have changed this issue through the text to clarify that this is not an objective of the classifications.

3- Type and necessity of normalisation depends on the type of data and the purpose of analysis or classification. If you compare runoff behaviour, normalization of data is necessary for each comparison of indices depending on catchment size. To compare runoff values of different catchments, normalization can be counterproductive. Therefore the aim of the classification determines normalization or not. For other indices like the timing of extreme flow events or numbers of days with increasing flow a normalization is meaningless. Please comment on this.

Referee was right on this comment and it is something we said in the introduction “normalization can be viewed as a completely subjective choice that depends on the purpose of the classification”. Nonetheless, we have rewritten the paragraph in the introduction to include some of ideas that were pointed out by the referee in order to clarify how normalization can affect classifications.

We also introduced several changes in section 4.2 of the discussion and in the conclusion to state clearly the main implications of normalize flow series for further uses of the classification.

4- Language: Frequently the text is difficult to read and imprecise. Many things remain unclear and should be revised.

We agree with the reviewer in that some of the sentences and paragraphs of the manuscript can be difficult to read. Two of the co-authors of the paper are English native and have made a rigorous language correction of the first version of the manuscript. Even tough, all the manuscript has been thoroughly revised and many sentences and paragraphs have been changed in order to clarify its actual meaning.

5- Why do you compare 19 classifications? Are there no optimal sizes of classification?

We compared 19 classifications because there was not an optimal number of classes that can be defined a priori. So in the paper we explored if according to the classification strength and the ANOVA analysis, the optimal number of classes for each classification procedure could be define. We observed that beyond 6-8 classes, the differences between classifications are not significant, i.e. classifications with different number of classes presented similar statistical performance. Hence, other criteria, different to the statistical performance, should be employed for the definition of the number of classes. We included an extra comment in the conclusion.

6- Why do you use a different number of hydrological indices (101 for the raw data against 103 for the normalized data)? May this affect the result?

When daily flow series were normalized by dividing each mean daily flow by the annual flow regime, l1 became equal to 1 in all the gauges. In addition, lcv became equal to l2 (as lcv = lca/l1). Hence, it makes no sense to include these variables in the analysis. For that reason when series were normalized the number of hydrological indices was reduced from 103 to 101. A brief explanation was now included in the new version of the manuscript.

7- Page 953, line 21: average rock hardness: which rock characteristic is the basis of the calculation and what is the meaning of the hardness to hydrological processes? Please explain.

A sentence explaining the base for the calculation of “rock hardness” and “permeability” has been included in the manuscript. Rock hardness affects significantly river morphology. River morphology interacts with the hydrological regime and influence, in part how the water flows through the reach (for instance, it can affect the duration of a high flow event). So that, we considered that it could be an interesting variable to include in the predictors (together with permeability), as there is little information regarding geology. However, it was less important than expected and in fact was one of the least important variables.

8- Page 957, line 5: acronym OBB unknown, or should this be OOB?

OBB was changed by OOB

9- Page 965: 4.3 Analyses of distinctive gauges belongs to results Fig. 5 and 6: unreadable small figures - perhaps better in another arrangement.

Reviewer was right with this comment. Figures 5 and 6 were included as graphical examples of the results of the ANOVA analysis obtained for all the hydrological indices (the results of each hydrological index were included in the supplementary material).. We reduced the number of variables presented in figures 5 and 6 and also enlarge all the symbols to increase its clarity and readability. We think that the reduction of the number of indices in the figures is still useful to understand the main results of the ANOVA.