

We thank the reviewer for his/her insightful comments. Below we respond to these comments.

“The authors did a nice job in setting up the modeling and the similarity frameworks, paying attention in defining a reasonably minimal parameterization compatible with accurate process description. I believe this is a very valuable step toward the goal of functional catchment classification”.

We thank the reviewer for the nice comments about our work.

“I would raise only very few points, trying to contribute to the clarification of the proposed approach. The first point attains the concept of ‘data’. The authors use hydrological signatures as catchment features that can be used to assess model performances. Two of the used signatures contain values that can unambiguously determined from observations, such as the runoff coefficient and the slope of the FDC. The third one, the Baseflow Index, is indeed an index that arises after filtering the streamflow data with a baseflow separation model. Even though much literature has been devoted to the identification of the baseflow index, I would suggest treating this quantity as a model-dependent variable, using caution in definitely considering it a ‘catchment feature’. A similar point has been treated by the authors in their statement of P4614 L9 ff, with reference to the characteristic time scales suggested in this paper. A correct distinction between actual data and ‘filtered’ data would imply a slight rephrasing of some parts of the paper in which the BI is implicitly treated as ‘data’”.

We agree that the computation of the baseflow index requires some sort of filtering of the streamflow data. However, we still think that it is correct to refer to this index as data-driven. The reason for this is that we simply pass the streamflow data through a low-pass filter, retaining only the low frequencies of hydrologic response and filtering out the high frequency response (rapid rise of the hydrograph related to quick flow). As hydrologists we tend to refer to this low frequency response as baseflow, but a more accurate term is slow flow. Thus the slow flow is derived from the data and model independent. The parameter ε controls the cut-off frequency and is not calibrated for each catchment (see also next comment) but kept constant at 0.925 across the 12 catchments (we forgot to mention this in the ms and this will be corrected in the revised version). Since the purpose of the study is to address hydrologic similarity across a climate gradient, the selection of a different cut-off level would not change the relative differences between the catchments (a desired characteristic of the data manipulation), but obviously will affect to some degree the absolute values. We would like to remind here that the baseflow index is compared to the modeled baseflow index by accounting for all subsurface flow generated by the model, not by separating the total streamflow generated by the model in a similar fashion as done with the streamflow data. This makes it a completely independent check on how the model is capable of correctly partitioning incoming energy and water fluxes into quick and slow flow components.

“Again with respect to the baseflow a more specific comment is in order. It looks not very clear how the authors calibrate the coefficient ε mentioned in eq. (33). I suppose

calibration would depend on an overall fitting between observed and modeled streamflow components (the time series or the Flow Duration Curves). Indeed, the role of ε is very significant, in that it determines the baseflow series from which the aquifer parameters are estimated (if I have correctly understood the procedure). If the technique for estimation of ε is based on a fitting of the FDC this would provide an additional explanation of the very good relation existing between the time scale of the deep aquifer and the slope of the FDC (P4612 L2 ff)”.

We apologize for not mentioning in the original manuscript that ε was kept constant across the catchments at a value of 0.925, in line with previous publications (e.g. Sivapalan *et al.*, WRR, 2011). Thus the parameter of the low-pass filter was not calibrated against data and therefore does not explain the good fit of our modeled slope of the flow duration curve. As mentioned in our ms, the goodness of fit is likely due to the calibration of the deep aquifer reservoir constant against observations condensed in the master recession curve.

“I find interesting the use the authors do of databases available for the catchments in the US. Of particular interest is the availability of soil information and of the parameter concerning vegetation height H . As this kind of data is not easily available in other geographic areas, it would be useful -in future work- to check the performance of methods that indirectly estimate H , which is a parameter of some relevance for the canopy losses. Incidentally, the same symbol, H , was already used for the sensible heat”.

Thanks again for your kind words. We initially select an average canopy height based on available land use data. However, as noted by this reviewer, this parameter controls the energy partitioning and we therefore included it in our parameter calibration procedure, as there is large uncertainty in determining a reasonable average canopy height in a mixed land use catchment. We will select a different symbol to indicate canopy height and reserve H for sensible heat (common notation).

“Finally, additional detail (e.g. a reference) would be appreciated as regards the University of Maryland vegetation classification system, that provides the fractional spatial coverage of vegetation type”

Thanks for catching this; we will provide references regarding this data set.

Sivapalan, M. *et al.*, 2011. Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry. *Water Resour. Res.*, 47(2), p.W02522.