**Hydrology and
Earth System
Sciences
Discussions**

Interactive comment on "Data-driven catchment classification: application to the PUB problem" *by* M. Di Prinzio et al.

M. Di Prinzio et al.

attilio.castellarin@unibo.it

We thankfully acknowledge Referee#2's comments, which we find extremely useful for improving the presentation of the work, enhancing its accuracy and rigorousness. We report below our replies (denoted by AR, Authors' Reply) to all referee's comments (indicated by RC).

RC:

General Comments

This is an interesting paper with a relevant contribution to the topic of catchment classiﬁcation. How data can be analyzed – in a meaningful manner – to understand hydrological similarity, and therefore allow catchment classiﬁcation is a current ques-

tion. Using SOMs in combination with data reduction techniques is one strategy to achieve this.

My main criticism at this point is that there is no attempt made by the authors for a physical interpretation (explanation?) of the result. The authors should expand their discussion to explain what catchments are grouped, why this might be hydrologically appropriate to group them (with respect to ﬂood behavior) etc. The authors present an interesting technique that they test in a reasonable way. Now they just have to expand it so that the paper becomes appropriate for a hydrology journal. This is something that I think the authors are very capable of doing.

AC:

We are a bit puzzled by Referee's main criticism. The main goal of the analysis is to assess whether (unsupervised and objective) multivariate techniques may improve the effectiveness of an unsupervised and objective approach (Self Organizing Maps, SOM's) to the problem of catchment classification within the PUB general context. SOM's count a number of hydrological applications presented in hydrological journals (see e.g. Hall and Minns, 1999; Jingyi and Hall, 2004; Srinivas et al., 2008; Toth, 2009). We show in the study qualitatively and quantitatively that SOM's benefit from the application of such multivariate techniques in predicting long-term streamflow indices (mean annual runoff) as well as flood flows.

If our main goal was to derive a physically-based catchment classification, assumed that this task can be carried out at all at national level in a country with such a huge hydrological variability, we would have probably adopted different approaches and techniques to begin with.

Our study falls instead within the vast literature applying multivariate statistical analyses, and in particular clustering algorithms based on geomorphologic and climatological catchment characteristics, to obtain an "objective" identification of groups (clusters) of watersheds with similar attributes. It is the information content of such attributes

that explains the similarity form a physical point of view. (See among the many others: Acreman and Sinclair, 1986; Wiltshire, 1986; Bhaskar and O'Connor, 1989; Burn, 1989; Nathan and McMahon, 1990; Burn et al., 1997; Burn and Goel, 2000; Madsen et al., 1997; Rao and Srinivas, 2008; Chawoshi and Soleiman , 2009; Guse et al., 2010).

We clearly point out in the discussion section and, in particular, in the Conclusions that "the application of objective but merely statistical criteria and algorithms (PCA and CCA with SOM) revealed some limitations that may be significantly reduced by switching from data-driven to data- and process-driven catchment classification".

Nevertheless, we believe that Referee's point may arise from a lack of clarity of our manuscript on the main goal of the study. We will revise the manuscript accordingly, by including a direct reference to the main goal outlined above in the introduction and also in the abstract, since we are asked to heavily restructure it (see below).

RC:

Specific Comments

- Please use more acronyms in your abstract!!! Just kidding. Please take out the acronyms there. It is not necessary and makes reading the abstract very cumbersome.

- It would be good to have less detail on the method in the abstract, but actually read about real results there so that the reader knows what he/she will get from reading the paper.

- The authors should avoid very short paragraphs. Certainly one sentence paragraphs are not good style.

- It would be good to deïňAne in the beginning of the paper (or in the abstract), that the authors are not talking about predictions of continuous streamïňĆow, but rather about different ïňĆow indices.

AC:

We will modify the abstract as recommended. Also, we will carefully revise abstract and body of the manuscript trying to make short sentences as long, verbose and unclear as possible!!! Just kidding.

We will avoid very short sentences in the revised manuscript.

RC:

- Is the ranking of controlling variables (i.e. controlling the classiïňAcation) similar?

- What assumptions are made regarding how the physical/climatic characteristics control the hydrologic behavior of catchments?

AC:

The Referee probably refers to the weight or importance of different descriptors for the identification of the different classes. For the same reasons reported in our reply to the main point raised by Referee#1, we preferred to refer in the study to the similarity and affinity of the different classes in terms of membership of different catchments (section 5.3) and to the performance of each classification for the prediction of streamflow indices (section 5.4).

To form the dataset of physiographic and climatic descriptors to be used in the study we did not make any a-priori assumption, since we were already constrained by the intrinsic difficulty of compiling a National dataset, which includes homogeneous and consistent information for ∼300 catchments. Therefore we included as many relevant catchment descriptors as possible, using multivariate analysis techniques to sort out noise and redundancy (Principal Component Analysis, see e.g. Chokmani and Ouarda, 2004; Castiglioni et al., 2011) while retaining the information that is more descriptive of the streamflow regime (Canonical Correlation Analysis, see e.g. Krzanowski, 1988; Ouarda et al., 2001).

Both these considerations were not explicitly reported in the text. We will include a reference to the first one at the beginning of section 5, which could also be useful

for introducing the structure of section 5 itself. We will also revise the manuscript by reporting a reference to our second comment at the beginning of subsection 4.1.

RC:

- I would separate section 2 into one section reviewing catchment classiïňĄcation and one discussing the issue of SOMs for classiïňĄcation in general.

AC:

Point taken. Section 2 will be revised according to Referee's comment, even though the main focus of the second subsection will necessarily be the hydrological application of SOM's rather than their application in general.

RC:

- It would be good to discuss (at the end) how this information (regionalized streamïňĆow indices) could be used further. For example, several authors (starting with Bardossy, 2007, JoH; and Yadav et al., 2007, Advances in Water Resources) suggest that these indices provide valuable information that can be assimilated into watershed models to reduce uncertainty in (continuous streamïňĆow) predictions in ungauged basins.

AC:

The revised manuscript will discuss the possible usefulness of catchment classification for the problem of rainfall-runoff model regionalization, including relevant references (we are aware of Bárdossy's study on regionalization of rainfall runoff models reported in HESS in 2007, but not of any published in JoH in 2007). The following paragraph will be included in the Introduction, around line 22 of page 393:

"In particular, catchment classification may support regionalization of rainfall-runoff parameters (Hundecha et al., 2008), a topical issue in hydrology (see e.g., Bardossy, 2007; Yadav et al., 2007; Castiglioni et al., 2010) which is also is particularly relevant

to the PUB-problem.

RC:

- It would be helpful to discuss how far this approach can be taken in the Italian context. For example, what other indices can (likely) be regionalized given what physical catchment descriptors are available? Mainly, what is known about subsurface characteristics.

AC:

Please refer to our reply to previous point on a National dataset of catchment descriptors and streamflow indices. We acknowledge that our dataset lacks on information concerning base-flow and subsurface characteristics in general, but also on land-cover and vegetation. Unfortunately we could not find consistent and homogeneous information on these characteristics nationwide. We will point out this limit of our study in section 4. of the revised manuscript (Study Area and available information).

RC:

- The mapping onto nodes within the SOM means that there is a frequency distribution 'in' each node. Could the uncertainty in this mapping be used to derive estimates of uncertainty in the predicted streamïňĆow indices?

AC:

This is a very good point. Even though this issue is clearly out of the scope of our study, we will cite this interesting research topic in the last paragraph of the discussion section (future analyses).

RC:

- It would be interesting if the authors would list the subjective choices that necessarily have to be made in this type of analysis, but that could inïňĆuence the outcome (e.g. the Euclidean distance measure). This might help to guide future studies.

AC:

We agree with the Referee on this point, and we will include a paragraph listing all subjective choices associated with implementation of SOM techniques (e.g., metric, network topology, number of classes, etc.) that may influence the application at the beginning of section 5.2 (SOM Classification and implementation, around L.6 on P.402).

REFERENCES

Acreman, M.C., C.D. Sinclair (1986) Classification of drainage basins according to their physical characteristics and application for flood frequency analysis in Scottland, Journal of Hydrology, 84(3), 365-380.

Bárdossy, A. (2007) Calibration of hydrological model parameters for ungauged catchments, Hydrol. Earth Syst. Sci., 11, 703–710, doi:10.5194/hess-11-703-2007.

Bhaskar, N. R., C.A. O'Connor (1989) Comparison of Method of Residuals and Cluster Analysis for Flood Regionalization, Journal of Water Resources Planning and Management, ASCE, 115(6).

Brath, A., A. Montanari, E. Toth (2002) Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models, Hydrology and Earth System Sciences, Vol. 6(4), pp. 627-640.

Burn, D.H., (1989) Cluster analysis as applied to regional flood frequency, Journal of Water Resources Planning and Management, 115(5), 567-582.

Burn, D.H. (1997) Catchment similarity for regional flood frequency analysis using seasonality measures. J. Hydrol. 202 (1–4), 212–230.

Burn, D.H., N.K. Goel (2000) The formation of groups for regional flood frequency analysis. Hydrolog. Sci. J. 45 (1), 97–112.

Castiglioni, S., A. Castellarin, A. Montanari, J.O. Skøien, G. Laaha, G. Blöschl (2011) Smooth regional estimation of low-flow indices: physiographical space based interpo-

lation and top-kriging, Hydrol. Earth Syst. Sci., 15, 715-727, www.hydrol-earth-syst-sci.net/15/715/2011/doi:10.5194/hess-15-715-2011

Castiglioni, S., L. Lombardi, E. Toth, A. Castellarin, A. Montanari (2010) Calibration of rainfall-runoff models in ungauged basins: a regional maximum likelihood approach, Adv. Water Resour., 33(10), 1235–1242, doi:10.1016/j.advwatres.2010.04.009.

Chavoshi, S., W.N.A. Solieiman (2009) Delineating Pooling Group for Flood Frequency Analysis Using Soft Computing, European Journal of Scientific Research, ISSN 1450-216X, 35(2), 181-187.

Chokmani, K., T. B: M. J. Ouarda (2004) Physiographical space-based kriging for regional ïňĆood frequency estimation at ungauged sites, Water Resour. Res., 40, W12514, doi:10.1029/2003WR002983.

Guse, B., A.H. Thieken, A. Castellarin, B.Merz (2010) Deriving probabilistic regional envelope curves with two pooling methods, Journal of Hydrology, Journal of Hydrology, Elsevier, 380, 14-26, 2010.

Hall, M.J., A.W. Minns (1999) The classification of hydrologically homogeneous regions. Hydrol. Sci. J. 44(5), 693–704.

Hundecha Y., T.B.M.J. Ouarda, A. Bárdossy (2008) Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the "spatial" structures of the parameters within a canonical physiographic-climatic space, Water Resour. Res. 44, W01427, doi:10.1029/2006WR005439.

Jingyi, Z. and M.J. Hall, (2004) Regional flood frequency analysis for the Gan-Ming River basin in China. J. Hydrol. 296, 98-117.

Krzanowski, W.J., Principles of Multivariate Analysis, Oxford University Press, Oxford, 1988.

Madsen, H., C.P. Pearson, D. Rosbjerg (1997) Comparison of annual maximum se-

ries and partial duration series methods for modelling extreme hydrological events 2. Regional modeling. Water Resour. Res. 33 (4), 759-769.

Nathan, R.J., T.A. McMahon (1990) Evaluation of Automated Techniques for Base Flow and Recession Analyses. Water Resources Research Vol. 26, No. 7, pp 1465-1473.

Ouarda, T., Girard, C., Cavadias, G.S. and Bobée, B. (2001) Regional flood frequency estimation with canonical correlation analysis, J. Hydrol., 254, 157–173.

Rao, A.R., V.V. Srinivas (eds.) (2008) Regionalization of Watersheds: An Approach Based on Cluster Analysis. Water Science and Technology Library, 58, Springer Science+Business Media B.V., New York.

Srinivas V.V., Tripathi S., Rao A.R., Govindaraju R.S., (2008) Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering, Journal of Hydrology, 348 (1-2), pp. 148-166.

Toth, E. (2009) Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting, Hydrology and Earth System Sciences, 13, 1555-1566.

Wiltshire, S.E. (1986) Identification of homogeneous regions for flood frequency analysis. J. Hydrol. 84 (3), 387-402.

Yadav, M., T. Wagener, H. Gupta, (2007) Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30(8), 1756–1774, doi:10.1016/j.advwatres.2007.01.005.