

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

Received and published: 13 April 2011

We thankfully acknowledge Referee#1's useful comments, which will significantly help us to improve the presentation of the study. We report below our replies (denoted by AR, Authors' Reply) to all referee's comments (indicated by RC).

RC:

The paper “Data-driven catchment classification: application to the PUB problem” investigates the issue of catchment classification through the use of unsupervised neural networks. In addition to the mere classification procedure, the approach is extended to allow regional estimation of some hydrological variables in ungauged basins. The paper is interesting and well written, and can be considered a valid contribution to

C823

the PUB issues.

A couple of general comments along with few minor notes are reported below. The SOMY classification used in the paper is based on the application of Self Organizing Maps to the whole set of hydrological attributes (six variables) as input layer. This leads to an “optimal” classification because it is based on all the available hydrological information; however, it would be interesting to look at the classification patterns also when only a subset of hydrological variables is considered. For instance, if one is interested on the classification based only on good-statistics, a possible subset of information could be the sample L moments (L_i with $i = 1 : 4$). This kind of classification could be different from the “global” one, and thus influencing the comparisons. Did you investigate the effect of different kind of hydrological inputs on the classification pattern?

AC:

The background idea is the identification of a catchment classification that could, in principle, serve different hydrological purposes and be used for addressing different PUB problems (for example design flood estimation or assessment of long-term surface water availability). This is the reason why we referred to a “global” classifications only, using Referee's wording. But the point is pertinent and relevant. We will modify the manuscript as follows:

(1) concerning the in Introduction, we will illustrate the background idea by including the above mentioned statement in the Introduction of the revised manuscript (around line 14 of page 394):

“The background idea is the identification of a multipurpose catchment classification that could, in principle, serve different hydrological analyses and be used for addressing different PUB problems (for example design flood estimation or assessment of long-term surface water availability).”

C824

(2) we will also modify the Conclusions by including in the last paragraph the following sentence:

“... open problem for future analyses. Our study focuses on a multipurpose catchment classification, future analyses will also consider hydrological classifications that are identified by focusing on a more specific water-problem, e.g. prediction of low-flows, flood flows, or surface water availability, to assess whether or not the same conclusions still hold.”

RC:

The authors conclude that an adequate number of classes, as a compromise between homogeneity and size of each class (page 402, lines 1-4), can be fixed equal to nine for the case study. The number of classes is the same for all the different classifications (SOMY and SOMX), and this hypothesis can certainly be considered adequate for exploratory analyses, such as this work is. However, the information content in the hydrological dataset Y is considered to be much “richer” than that of the descriptor dataset X. This gap is well known, in fact, there is a general claim for new types of descriptor that include process-related information, in order to improve the classification/regionalization procedures when using only non-hydrological descriptors. The point is: if X contains less useful information than Y, it is probably sufficient a lower number of (larger) classes. Did you evaluate any scenarios with a different number of classes between SOMY and SOMX during the exploratory analysis?

AC:

Exploratory analyses (see Toth & Castellarin, 2008) revealed that 9 classes is a good compromise between homogeneity and size of each class for classes identified on the basis of the physiographic and climatic descriptors (12 X Variables in this study). Homogeneity has to be interpreted in terms of variability of physiographic and climatic catchment descriptors in this context. Therefore, 9 was selected as the number of classes for all classifications in order for the comparison to be consistent.

C825

Some classifications are based on a much smaller number of descriptors (e.g. 3 for SOMPC3). Concerning this point, even though 3 principal components explain the larger amount of variability in the original set of descriptors through linear combinations, it has to be remembered that SOMs are formed through a non-linear and unsupervised process. We therefore believe that a larger number of classes may be adequate also in this case and deemed 9 classes suitable for all classifications (the ones based upon 6 Y variables, and those based upon X variables and their linear combinations).

In addition, in the present exploratory work we preferred to preserve the same number of classes for all methods to facilitate the comparison. In particular, the original Rand Index is sensitive to the number of classes in the partitions, and may be significantly impacted (i.e. one obtains lower index values) when the considered classification has a different number of classes from the reference one. We acknowledge that these points were unclear in the original version of the manuscript, and we will clarify them in section 5.2.

RC:

Line 6 page 396: I would briefly explain the meaning of “preserve the topology” in the text.

AC:

According to Referee's suggestion, we will revise the text by explaining what we meant by “preserve the topology”. In particular, we will include the following statement in section 2. (around L.17 on P. 396): “Lateral interaction between neighbouring output nodes ensures that learning is a topology-preserving process in which the network adapts to respond in different locations of the output layer for inputs that differ, while similar input patterns activate units that are close together.”

RC:

Section 5.2: I would report some information about the typical size of each class and

C826

the variability of the between-classes number of elements.

AC:

We will include either a table or a graphical representation of the number of member of each class for all of the identified classifications.

RC:

Technical notes

Throughout the text: I am not a native English speaker, but I found many clues about the fact that the plural acronyms should be written without the apostrophe (e.g. SOMs instead of SOM's). Please verify that.

AC:

We believe that both styles are used, including the apostrophe is the traditional way (see e.g. http://en.wikipedia.org/wiki/Acronym_and_initialism#Representing_plurals_and_possessives), we are open to use any of the two, following the indications of the Journal production office.

RC:

Line 6 page 397: remove one "PCA" duplicated. Table 1: Missing ")" at the end of the caption. Table 5. I suggest to move "minimum record length" to the table header and put "none" or "no limit" in the MAR and I1 lines.

AC:

Many thanks, the error will be corrected and suggestions incorporated.

REFERENCES

Toth, Castellarin (2008) Catchment classification coupling Canonical Correlation Analysis and Self Organising Maps, Geophysical Research Abstracts, 10, EGU2008-C827

A-12015

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 391, 2011.