

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

Anonymous Referee #1

Received and published: 9 March 2011

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 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

C368

also when only a subset of hydrological variables is considered. For instance, if one is interested on the classification based only on flood-statistics, a possible subset of information could be the sample L moments (L_i with $i = 1 \dots 4$). 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?

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 considered adequate for exploratory analyses, such 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?

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

Section 5.2: I would report some information about the typical size of each class and the variability of the between-classes number of elements.

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.

C369

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.

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