Hydrol. Earth Syst. Sci. Discuss., 9, C5609–C5616, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C5609/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Catchment classification based on characterisation of streamflow and precipitation time-series" *by* E. Toth

E. Toth

elena.toth@unibo.it

Received and published: 30 November 2012

I wish to deeply thank Alberto Viglione for having so carefully read the paper and provided useful insights; his comments are particularly valuable given his deep knowledge and expertise in catchment classification and applications to the PUB problem.

REPLY TO ALBERTO VIGLIONE'S COMMENTS

"1) The paper is about classifying based on hydrological similarity, which is more general than similarity in terms of floods, or seasonality, or low-flows, or other singular hydrological characteristics. How would the author define this more general similarity?"

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

R. Actually, this is a very difficult question, maybe a sort of streamflow regime similarity, using ‘regime’ in a nonspecific, broad sense referring to a sort of ‘predisposition’ of the catchment behaviour? Obviously the ambition would be to be able to represent all aspects of the hydrologic behaviour, but – even setting aside the limitations highlighted in the manuscript - this kind of approach can not be able to capture the catchment response in correspondence of extreme events. On the other hand, it may be useful for obtaining indications for choosing the structure and parameterisation of a rainfall-runoff model to be used for continuous simulation.

“2) As a reference, the author assumes that the clusters obtained based on many runoff statistics are indeed hydrologically similar. Why have these statistics been chosen and not others? Would the neural network procedure give the same results if the streamflow signatures scaled by the mean would be used (see below)?”

Comment 2) is related to Comments 8) and 16):

“8) Page 10814, lines 12-21: the trained network mainly distinguishes between catchments with low and high runoff (dry and wet), while the lag-1 autocorrelation coefficient and the correlation scaling exponent have not high discriminant power. My feeling is that this is because the other runoff signatures taken into account (mu_Q, P_Q5, P_Q95 and sigma_Q) all are expressed in mm/h, i.e., all represent a volume of runoff. I wonder what would be the result if mu_Q, P_Q5/mu_Q, P_Q95/mu_Q and CV_Q=sigma_Q/mu_Q were used. Would it make a difference?”

“16) Page 10819, lines 7-16: as noted before, the fact that the dynamic component of the streamflow does not play a major role in the classification may be due to the fact that all other signatures are volumetric, therefore much more weight is given to the wetness of the catchment. Or is the redundancy of information properly accounted for in the neural network technique?”

R. Of course different statistics representing the streamflow regime might have been selected, included a scaling or any other combination of the chosen ones. But, actually, in order to investigate the significance of the information content of the different variables, I had initially carried out a Principal Component Analysis also on the streamflow signatures set, and not only on the catchment descriptors.

The first three PCs explains more than 90% of the variance and a SOM applied providing in input the vectors formed by the first 3 PCs of the streamflow signatures led to clusters (see Figure 1 attached below, 'Classification identified by SOM based on the first three PCs of the streamflow signatures') that are identical to those obtained based on the original streamflow signature vectors (Fig. 1 of the discussion manuscript).

Since the PCs should not overlap in information content, the fact that the original and the PCs streamflow variables lead to the same classification seems to indicate that the use of the not scaled indexes does not suffer of redundancy of information.

In the paper, since the above results demonstrate that a dimension reduction of the streamflow attributes was not necessary in the SOM application (whereas it was necessary in the case of the catchment attributes, since for applying the Discriminant Analysis, see Comment 11), I preferred not even citing the PC Analysis of the streamflow signatures, in order to not further confuse the Readers with another PCA (as evidenced by both Ref #1 and #3 the unsupervised (SOM for streamflow attributes)/supervised (PCA+Discr. Analysis for catchment attributes) procedure is already pretty complicated to follow. . .)

In addition, also the SOM may be used as an information content extractor (and, if needed, as a dimension reduction technique), projecting the analysed entities on the output layer which has a lower (2) dimension than that of the input vectors. In the original manuscript the number of nodes was limited to 3, but, following Ref. 1's suggestion, I intend to analyse the results obtainable with a larger output layer (corresponding to more classes and therefore to a more detailed classification) hoping that such analysis may help to better understand the role/limitations of the proposed dynamic component signatures.

HESSD
9, C5609–C5616, 2012

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

“3) The neural network procedure is used for streamflow statistics and the PCA+discriminant analysis for catchment/climate attributes. Why not using the PCA+discriminant analysis technique for both?”

This Comment is related to Comments 13) and 14):

“13) Page 10817, lines 14-27: Wouldn’t it be possible to cluster the catchments based on catchment attributes using SOM? Is it because an allocation rule for ungauged catchments cannot be defined through SOM?”

“14) Page 10817, line 25: "...and the classes are the three clusters identified by the SOM network based on the streamflow signatures", does the discriminant analysis with catchment characteristics take the 3 classes obtained with SOM on streamflow characteristics as an input? I’m confused.”

R. As I wrote in the Reply to Referee #1, it is evident that the paper is not clear in the description of the unsupervised/supervised approach and I will try to better describe these steps in the revised paper, hoping that the methodology may result clearer. The PCA+ Discriminant Analysis approach can not be applied for classifying the streamflow indexes, because the Discriminant Analysis is a supervised learning technique, that assigns each record to predefined groups, that here are not available. (Comment 3) For this reason the SOM is first applied as an unsupervised methodology for grouping catchments that are similar from the hydrometric point of view. Successively, aiming at assigning to such classes also any new watershed where hydrometric measures are not available, the Discriminant Analysis is applied (in the leave-one-out validation framework), constructing a classification rule based on the knowledge of i) the morpho-pluviometric attributes and ii) the hydrometric classification of the gauged watersheds (such classification, obtained by the unsupervised SOM based on streamflow indexes, becomes, in this second analysis, our predefined grouping, our reference). In this sense, the Discriminant Analysis needs to know the SOM-classes, to identify a rule that associates, for each gauged catchment, the set of catchment attributes to the corresponding (streamflow-based) class; such rule is then used to associate to one of the three streamflow-based classes the ungauged basins as a function of their catchment attributes. (Comment 14)

As far as Comment 13) is concerned: a SOM may be used for carrying out a new, different unsupervised classification, based on catchment attributes alone (see previous analyses such as Di Prinzio et al, 2011), but not for assigning a new ungauged entity to a predetermined group, since the SOM input must be a vector of streamflow indexes, that are not available for ungauged basins.

“4) Page 10806, lines 19-23 (Abstract): I would suggest to substitute "quite satisfactory results" and "acceptable overlap" with more quantitative measures, e.g., misclassification rate (20%).”

R. Thank you for the suggestion; I will modify the text accordingly.

“5) Page 10808, line 22-23: The following sentence is unclear to me: "Since the time-series autocorrelation functions might differ strongly one from another, their comparison and classification may be extremely difficult". If the autocorrelation functions differ strongly one from another, I would conclude that the catchments are different in terms of storages. Or, does the author mean that sample autocorrelation estimators are not robust and therefore other authors use parametric methods (parameters of a linear model)?”

R. The wording of the phrase is probably confusing: I simply meant that the shapes of the autocorrelograms may be very different and that a comparison through a visual inspection or a synthesising index may not be easy. I will rephrase it to make it understandable.

“6) Page 10811, line 11: remove "obtained".”

“7) Page 10812, line 7: I guess it is "high resolution areal rainfall time-series".”

R. You are absolutely right! I will correct the errors. . .

HESSD

9, C5609–C5616, 2012

Interactive Comment

8) see above

“9) Page 10815, line 1: how high are the western mountains?”

R. The highest mountains of the Emilia (western) region are more than 2000 m.a.s.l.

“10) Page 10815, line 5: what is the annual rainfall depth in the Romagna area close to the sea?”

R. On the coast from 650 to 700 mm; our watersheds have higher altitude and the rainfall depths are of course higher (for the three more eastern watersheds: from 840 to 1015 mm per year in the observation period).

“11) Page 10816, line 10: I do not understand clearly the sentence with "it follows that", how comes that no more than 3 or 4 discriminant variables should be used? (Maybe it’s just me)”

R. Having 44 catchments as a whole, to be divided in 3 clusters, we have (44 divided by 3, that is) 14 or 15 entities for each cluster on average; such number, divided by 4 (the number of entities needed for each discriminant quantitative variable for each cluster) gives us the maximum number of discriminant quantitative variables, that is between 3 and 4. Or, going backward, 4 entities (catchments) multiplied by 3 or 4 discriminant variables, multiplied by 3 clusters gives us approximately the size of our data set.

“12) Page 10817, line 10: I would suggest "...is more temporally correlated than..." ”

R. Thank you for the suggestion; I will modify the text accordingly.

13) see above

14) see above

“15) Page 10818, lines 6-17: maybe recall here that class 1 is the dryer/low elevated and class 3 the wetter/high elevated”

R. Thank you for the suggestion; I will modify the text accordingly.

16) see above

“17) Figure 1 and 2: as pointed out by another reviewer, maybe some more information could be added to the figures, such as the topography (in one) and the mean annual precipitation (in the other). Based on the discussions in the paper, both should be very much correlated to the catchment grouping.”

R. As written in the reply to Referee #2 I fully agree that adding i) information on the variability of the catchment attributes (either with additional maps or including it in Fig. 1 and 2) and ii) some details on the interpretation of the results would improve the understanding of the paper.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 10805, 2012.

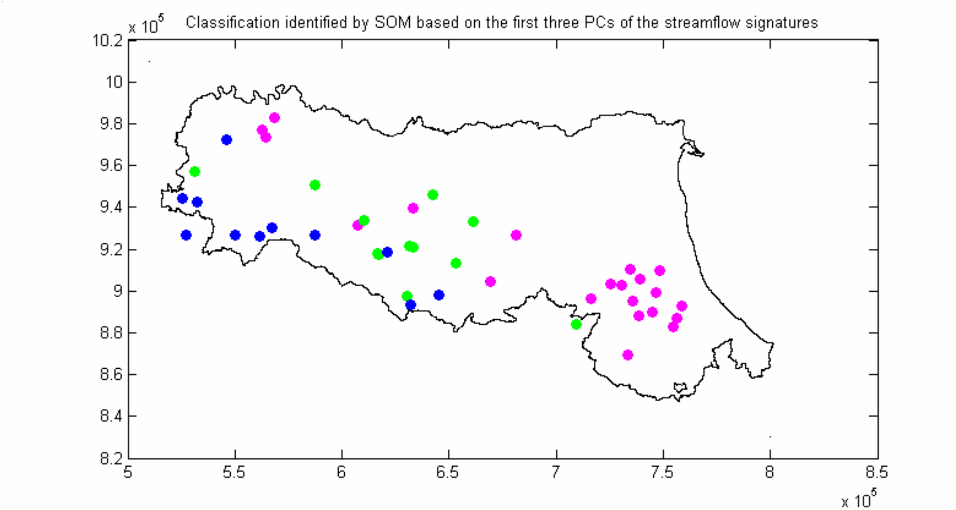

Fig. 1. Classification identified by SOM based on the first three PCs of the streamflow signatures

HESSD
9, C5609–C5616, 2012

Interactive Comment

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper