Hydrol. Earth Syst. Sci. Discuss., 9, C5103–C5106, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C5103/2012/

Hydrology and Earth System Sciences Discussions

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

A. Viglione (Referee)

viglione@hydro.tuwien.ac.at

Received and published: 3 November 2012

In the paper two automatic classification methods are used to group catchments: 1) a neural network method with streamflow statistics, and 2) a principal component + discriminant analysis with catchment/climate attributes. Based on a cross-validation procedure, the paper shows that the method based on catchment/climate attributes assigns ungauged catchments to the same groups obtained based on streamflow statistics 80% of the time. This is indeed a very promising result.

I enjoyed reading the paper, which is well written and to the point. Therefore I am supportive for the publication of the paper. However I believe that some more discussion on the (minor) comments listed below would be useful.

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?

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)?

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?

Detailed comments:

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

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)?

Page 10811, line 11: remove "obtained".

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

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?

Page 10815, line 1: how high are the western mountains?

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

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)

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

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?

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.

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

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 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 acconted for in the neural network technique?

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.

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