Hydrol. Earth Syst. Sci. Discuss., 9, 10805–10828, 2012
www.hydrol-earth-syst-sci-discuss.net/9/10805/2012/
doi:10.5194/hessd-9-10805-2012

Hydrology and Earth System Sciences Discussions

Catchment classification based on characterisation of streamflow and precipitation time-series

E. Toth

Department DICAM, University of Bologna, Italy

Received: 24 August 2012 – Accepted: 11 September 2012 – Published: 26 September 2012

Correspondence to: E. Toth (elena.toth@unibo.it)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The formulation of objective procedures for the delineation of homogeneous groups of catchments is a fundamental issue in both operational and research hydrology. For assessing catchment similarity, a variety of hydrological information may be considered; in this paper, gauged sites are characterised by a set of streamflow signatures that include a representation, albeit simplified, of the properties of fine time-scale flow series and in particular of the dynamic components of the data, in order to keep into account the sequential order and the stochastic nature of the streamflow process. The streamflow signatures are provided in input to a clustering algorithm based on unsupervised SOM neural networks, providing an overall reasonable grouping of catchments on the basis of their hydrological response. In order to assign ungauged sites to such groups, the catchments are represented through a parsimonious set of morphometric and pluviometric variables, including also indexes that attempt to synthesize the variability and correlation properties of the precipitation time-series, thus providing information on the type of weather forcing that is specific to each basin. Following a principal components analysis, needed for synthesizing and better understanding the morpho-pluviometric catchment properties, a discriminant analysis finally classifies the ungauged catchments, through a leave-one-out cross-validation, to one of the above identified hydrologic response classes. The approach delivers quite satisfactory results for ungauged catchments, since the comparison of the two cluster sets shows an acceptable overlap. Overall results indicate that the inclusion of information on the properties of the fine time-scale streamflow and rainfall time-series may be a promising way for better representing the hydrologic and climatic character of the study catchments.

HESSD

9, 10805–10828, 2012

1 Introduction

The identification of groups of hydrologically similar catchments is a fundamental issue in both operational and research hydrology: it is essential to ensure the transferability of information when applying regionalization methods, but can also provide valuable indications to improve the understanding of the dominant physical phenomena in the different groups (McDonnell and Woods, 2004; Wagener et al., 2007; Sawicz et al., 2011). The similarity may be evaluated in terms of signatures of catchments' functional responses, quantifying the characteristics of the hydrologic response that provide insight into the behavior of the catchment (Atkinson et al., 2002; Wagener et al., 2007; Yilmaz et al., 2008; Oudin et al., 2010). A comprehensive set of measures describing all aspects of the catchment hydrology (such as meteorological observations, soil moisture content, vegetation patterns, etc.) should in principle be analyzed in order to fully understand these functional characteristics, but unfortunately such measures are not available in the majority of catchments. It is therefore worthy analyzing the information content embedded in data far more generally available, such as streamflow measures, even if acknowledging that in this way, while it is possible to include in the study a much greater number of catchments, the similarity analysis can provide only a first-order classification (Wagener et al., 2007; Sawicz et al., 2011). On the other hand, streamflow may be seen as an integrator of all climatic and morphologic conditions of a given basin (Samaniego et al., 2010), thus justifying such empirical approach. To this end, a variety of indexes based on streamflow measurements may be adopted, characterizing in a different way the hydrological response of the basin, generally depending on the type of analysis to be carried out. The most frequent and compelling need for the assessment of regional similarity in catchment response is in fact for issuing predictions in ungauged catchments and the choice of the streamflow indexes to be compared depends on the finality of the regional analysis, that is on the variable to be predicted. The large majority of regionalisation studies performing an objective catchment classification, through the use of clustering techniques, has concerned, since

HESSD
9, 10805–10828, 2012

the 80's, flood frequency analysis (e.g. Hosking et al., 1985; Lettenmaier et al., 1987; Burn, 1989; Burn et al., 1997; Burn and Goel, 2000; Castellarin et al., 2001; Merz and Bloeschl, 2005). For such analyses, the main representative streamflow variables are, naturally, the flood peaks values. If the objective is, instead, the assessment of water availability, the streamflow indexes to be predicted may be for example mean annual or monthly flows (e.g. Haines et al., 1988; Holmes et al., 1999) or low flow percentiles (e.g. Nathan and McMahon, 1990; Laaha and Bloeschl, 2006; Vezza et al., 2010) or the entire flow duration curve (e.g. Singh et al., 2001; Ley et al., 2011; Patil et al., 2011; Sauquet and Catalogne, 2011). On the other hand, such representations do not allow to keep into account the sequential order and the stochastic nature of the streamflow process; these properties would, for example, be crucial if the regionalization aimed, as often needed in the hydrological practice, at the parameterization of a rainfall-runoff model at fine temporal scale and the catchment similarity should therefore be guaranteed in terms of continuous streamflow generation. It may therefore be important also representing and comparing, in addition to mean values or percentiles, the properties of the low time-scale streamflow series and in particular the dynamic components of the data. Information on the effect of complex driving factors on the hydrological response (not always easy to recognize) are in fact embedded in the temporal dynamics of the streamflow series (Chiang et al., 2002; Corduas, 2011). Important differences among the streamflow processes may be highlighted by the analysis of their temporal correlation structure, representable through the global autocorrelation function ACF (or the corresponding power spectrum). Since the time-series autocorrelation functions might differ strongly one from another, their comparison and classification may be extremely difficult. To tackle this issue, recent studies (De Thomasis and Grimaldi, 2001; Chiang et al., 2002; Corduas, 2011) proposed to analyze the streamflow temporal dynamics through the parameter sets of linear models estimated on the corresponding streamflow time-series. A more parsimonious, but less refined and necessarily approximated, approach is here applied for representing the autocorrelation structure: in addition to the lag-1 autocorrelation coefficient (previously used in regionalization studies, for example, by Montanari and Toth (2007); Castiglioni et al. (2010); Lombardi et al. (2012) for the parameterization of a rainfall-runoff model), it is here proposed to use an index representing the shape of the ACF, that is the *correlation scaling exponent*, which has been used for analysing the scale properties of meteorological and hydrological data (see, e.g. Menabde et al., 1997; Marani, 2003; Molnar and Burlando, 2008; Ozger et al., 2012).

Section 2 presents the study area and the indexes estimated for both gauged and ungauged catchments; in Sect. 3, the set of descriptors summarising the main statistical features of the streamflow time-series (including the coefficients above cited for representing the temporal correlation structure) are provided in input to a clustering algorithm based on unsupervised SOM neural networks, recently proposed for catchment classification, but so far never utilised for classifying attributes based on time-series properties.

The final aim of the study is the assignment of ungauged catchments to the classes obtained from the similarity of the river flow time series, presented in Sect. 4. To this end, the ungauged catchments are characterised through a set of indexes describing their morphology and the main rainfall properties. In particular, besides the morphological indices, it was deemed appropriate to rely on the information content of long, high-resolution rainfall time-series. In analogy with the streamflow series represntation, the rainfall attributes include also the indexes describing the temporal variability of the series, that allow to incorporate information on the dynamics of the process, thus characterizing the type of weather forcing that is specific to each basin. The present paper provides the first ever catchment classification to be performed including coefficients (and in particular the correlation scaling one) characterising the fine time-scale variability and correlation structure of both streamflow and rainfall fine-resolution time-series. Following a principal components analysis (Sect. 4.1), needed for synthesising and better understanding the morpho-pluviometric catchment properties, a discriminant analysis (Sect. 4.2) is then applied in a leave-one-out cross validation approach, to identify the membership of ungauged catchments to the original hydrometric classes.

HESSD
9, 10805–10828, 2012

It is therefore finally possible to determine the error rate for classifying the streamflow properties based on catchment descriptors that are available also in absence of hydrometric measurements.

2 Study area and classification attributes

2.1 Study area and data

The study region includes 44 catchments, spanning the north-eastern side of the Apennine mountains and piedmont area (Emilia-Romagna), in Italy. The Apennines of North-Central Italy are a fold-and thrust mountain chain related to an orogenic system (chain-foredeep-foreland), derived from the post-Eocene collisional history between the European and African plates and from a complex, multi-staged evolution. The topographic relief is made up of a series of ridges elongated in directions that vary from S–N to SW–NE, separated one from the other by narrow valleys or by wide intermontane tectonic depressions (Piacentini et al., 2011). The landscape is rougher and steeper in the western chain, whereas the Adriatic piedmont areas are characterized mostly by gently reliefs down to the coastal lowlands.

For each of the study catchments, time-series data of hourly streamflow were collected for a total number of observations ranging, for the different river sections, from 31 519 to 85 469 (that would correspond respectively to more than 3.5 and almost 10 yr of continuous monitoring but, as a matter of fact, embrace periods of missing data). Hourly streamflow data are expressed as spatial averaged runoff depths (mm h^{-1}). Areal precipitation estimates, again at hourly step (mm h^{-1}), were interpolated with Thyssen-polygon weighting from nearby rain-gauges.

2.2 Streamflow signatures

The first step of the proposed approach is to cluster the catchments on the basis of the hydrologic response, as defined by key signatures of the streamflow time-series.

HESSD
9, 10805–10828, 2012

The chosen signatures are: (i) average runoff, μ_Q, (ii)–(iii) the 5th and 95th percentiles, $P_{Q,5}$ and $P_{Q,95}$ and (iv) the standard deviation, σ_Q, of hourly data. To describe the correlation structure of the series, representing the dynamic component of the process, two metrics were computed: (v) the lag-1 autocorrelation coefficient, $\rho_Q(1)$, and (vi) the *correlation scaling exponent* (see for example Menabde et al., 1997, and Molnar and Burlando, 2008, for precipitation data, and the recent application by Ozger et al., 2012, to streamflow time-series), that is the exponent that characterises the correlation function with a power law:

$$\rho_Q(\tau) \propto \tau^{-\alpha_Q} \tag{1}$$

where ρ_Q is the autocorrelation function, τ is the time lag, and α_Q is the correlation scaling exponent, obtained. Values of α_Q tending to 0 indicate strongly correlated data, values close or higher than 1 show absence of correlation. Actually, the analysis of the correlation structures would require stationary time-series, whereas streamflow observations (as well as rainfall ones) exhibit a strong dependency on the season (see also the recent analysis by Patil et al., 2011); to solve this problem, the above cited papers assume stationarity on a seasonal basis, estimating separate coefficients for the different seasons or months. In addition, in the presence of trends, Eq. (1) may not be capable of characterizing the structure of data in terms of multifractality and correlation dimension (see, e.g. Ozger et al., 2012). Nonetheless, due to the limited number of catchments in the data set, it was deemed appropriate, in this first study, to retain the smallest possible number of streamflow signatures, for avoiding over-parameterisation effects in the classification technique; for this reason, even if acknowledging the strong limitations of this approximation, only one value for α_Q was estimated for each time-series.

2.3 Catchment descriptors

In order to extend the analysis of the hydrological similarity also to catchments devoid of flow measurements, indexes describing the basins from the geo-morphological and

HESSD
9, 10805–10828, 2012

climatological point of view are identified. The main geographical and morphometric attributes are derived from digital catchment boundaries coupled with the digital elevation model: (i)–(ii) the geographical coordinates UTM X and Y of the streamgauges; (iii) drainage area, A, (iv)–(v) minimum and average catchment elevation, H_{min} and H_{med}, and (vi) main stream length, L. In addition, to better describe the catchments as far as the rainfall-runoff transformation is concerned, indexes obtained from the low-resolution areal rainfall time-series are estimated, thus attempting to characterise the fine time-scale variability and correlation structure of the precipitation process. The chosen pluviometric attributes are: (i)–(ii) the mean and the standard deviation of the hourly data, μ_P and σ_P; (iii) the average proportion of wet hours (hours with more than 0.2 mm of rain), P_{Wet}; finally, in analogy with the streamflow signatures, (iv) the lag-1 autocorrelation coefficient, $\rho_P(1)$, and (v) the correlation scaling exponent, α_P, of the precipitation time-series are computed.

The chosen streamflow signatures and catchment attributes (pluviometric and morphometric) are listed in Table 1, along with the corresponding observation ranges over the data-set.

3 Classification of streamflow signatures with SOM neural networks

In the past three decades a number of applications of cluster analysis techniques have been presented in the hydrologic literature for the objective identification of catchments having similar attributes (either geographic, morphometric, climatic and/or based on streamflow observations). In the recent years, also non supervised neural networks, and in particular of the SOM (Self-Organising Mapping) type, were successfully applied for catchments classification purposes (Hall and Minns, 1999; Hall et al., 2002; Jingyi and Hall, 2004; Srinivas et al., 2008; Di Prinzio et al., 2011; Ley et al., 2011). SOM-type neural networks learn to cluster the input data by recognizing different patterns organising the data on the basis of their similarity, quantified by means of a distance measure

HESSD
9, 10805–10828, 2012

(in the present case, like in the majority of applications, the Euclidean distance). More details on the SOMs and in particular on their use as classification techniques may be found for example in Herbst and Casper (2008) or in Toth (2009). The networks are formed by two layers of interconnected nodes (or neurons): each attribute of the entity to be classified (i.e. a catchment) is fed to one of the input nodes, while the output nodes correspond to the classes to which the entities are assigned. An input vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ activates in fact only one output node, representing its class, using the Kohonen competitive learning rule (Kohonen, 1997). Each output node is characterized by the weights connecting it to the input nodes. Initially the weights between the n input nodes and each output node are randomly assigned. When, in the training phase, an input is sent through the network, each output neuron computes the distance between its weights $W = (w_1, w_2, \ldots, w_n)$ and the input vector:

$$\|\boldsymbol{x} - W\| = \sqrt{\sum_{i=1}^{n} (x_i - w_i)^2} \tag{2}$$

The output node responding maximally to the given input vector – that is the weights vector having the minimum distance from the input vector – is the winning neuron. At each training iteration t, the weights of the winning node and of its neighbouring nodes change, so to further reduce the distance between the weights and the input vector:

$$W(t+1) = W(t) + \mu(t) h_{lm} (\boldsymbol{x} - W(t)) \tag{3}$$

where μ is the learning rate, $\in[0\ 1]$, $\boldsymbol{l}$ and $\boldsymbol{m}$ are the positions of the winning and its neighbouring output nodes and h_{lm} is the neighbourhood shape, that reduces the adjustment for increasing distance, namely,

$$h_{lm} = \exp\left(-\frac{\|\boldsymbol{l} - \boldsymbol{m}\|^2}{2\theta(t)^2}\right). \tag{4}$$

HESSD
9, 10805–10828, 2012

where $\|\boldsymbol{l} - \boldsymbol{m}\|$ is the lateral distance between $\boldsymbol{l}$ and $\boldsymbol{m}$ on the output grid and θ is the width of the topological neighbourhood. 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 adjacent output units, corresponding to akin classes.

The values of the six streamflow signatures (i.e. the input vectors) are standardized to zero mean and unit variance, so to give them equal importance in the evaluation of the distance measure. To avoid the effects of over-parameterization, given the total number of gauged catchments, it was decided to limit to three the number of hydrologically homogeneous clusters. The network therefore consists of an input layer and an output layer formed by 6 and 3 nodes, respectively.

When the training is complete, each vector of streamflow signatures is assigned to its winning node, that corresponds to the class. The trained network associates to class 1 almost half of the study catchments (21 over 44), corresponding to those that generate the lower runoff (smaller values of μ_Q, $P_{Q,5}$, $P_{Q,95}$ and σ_Q) and, on the whole, less strongly autocorrelated series (smaller $\rho_Q(1)$ and greater α_Q coefficients, on average). classes 2 and 3 (formed by 12 and 11 elements respectively), include catchments generating higher runoff (greater for class 3) and with higher temporal correlation, on average. It must be highlighted, however, that the values of $\rho_Q(1)$ and α_Q are extremely variable inside each class, thus suggesting to test, in future investigations, also different indexes for representing the dynamic component of the process.

Figure 1 shows the closure sections of the catchments associated to the three clusters obtained from the streamflow signatures. It may be observed that class 1 includes almost all the basins of the south-eastern part of the study area and some western sections located in the lower part of the valleys. The south-eastern part of the region (named *Romagna*) is actually a different hydrographic region, since it is formed by rivers flowing directly in the Adriatic Sea, while the remaining catchments are all headwater tributaries to the Po River (the most important Italian river), belonging to the western part of the region (*Emilia*). Also the climate varies between the two areas

HESSD
9, 10805–10828, 2012

Catchment classification

E. Toth

from mountainous to maritime, going from the higher crests of the western side to the eastern coastal and hilly area. The western side of the region experiences more rain with annual rainfall depths that exceed 2000 mm in the mountains, whereas the climate in the Romagna area changes due to the wind exposition, to the influence of the sea, to the lower orography and also to the lower latitude. It is realistic that the Romagna catchments, that are close to each other, behave in a hydrologically similar way, since contiguous areas are characterized by similar climate, topography and geology, and all other characteristics deriving from them, such as soil type, vegetation, etc. (Merz and Bloeschl, 2005; Patil et al., 2012). Along with the Romagna catchments, class 1 includes other lower-altitude and less rainy catchments. To class 3 mainly belong, instead, the mountainous catchments of the western area, characterised by higher altitude and smaller drainage areas; finally, class 2 stands in between the other two classes, with principally western piedmont stations. The SOM classification based on streamflow signatures seems therefore to indicate, overall, a reasonable grouping ability, highlighted by the spatial distribution of the classes of homogeneous catchments. In fact, even if no geographical nor elevation information is provided in input to the SOM network, the clusters exhibit an overall consistency as far as location and altitude are concerned.

4 Application to ungauged catchments

One of the primary practical objectives for delineating hydrological homogeneous regions is to assess the membership of ungauged sites, thus inferring indications on the response behaviour of such catchments. An important feature of a cluster analysis aimed at identifying homogeneous clusters is therefore the ability to discriminate between them on the basis of variables that are different from the streamflow signatures, that is, a set of physical and climatic characteristics of the watersheds. To this end, a discriminant analysis will be applied to predict the classes – as identified in the previous section – of ungauged drainage basins described by means of the morphologic

HESSD
9, 10805–10828, 2012

and pluviometric properties presented in Sect. 2.3, chosen as discriminant. This approach is similar to that applied by Bhaskar and O'Connor (1989), Chiang et al. (2002) and Sanborn and Bledsoe (2006), albeit with different hydrometric and morpho-climatic sets of indices and different cluster analysis techniques.

4.1 Principal component analysis of catchment descriptors

The chosen morphometric and pluviometric catchment descriptors available for ungauged stream-sections are a total of 11 (X, Y, A, $H_{\min}$, H_{med}, L, μ_P, σ_P, P_{Wet}, $\rho_P(1)$, α_P). For an optimal discriminant analysis it is preferable (Hand, 1997; Sanborn and Bledsoe, 2006) to have at least four entities (catchments) for each discriminant quantitative variable for each cluster; it follows that, since the streamflow signatures classification has divided 44 catchments in three clusters, it is advisable to use a set of no more than 3 or 4 discriminant variables. The catchment descriptors vectors are therefore subjected to a principal component analysis to identify a smaller number of uncorrelated variables that describe the dominant patterns of variance in the data.

The principal component analysis shows how the first three principal components (PCs) explain, together, 89 % of the total variance (and the fourth one adds only another 4.6 % of explanation). It is therefore chosen to synthesise the information content of the 11 morpho-pluviometric variables through the first three PCs. Table 2 presents the original variables that mostly affect the first three PCs (that is, those with the highest loadings) and the percentage of total variance explained by each PC.

The principal component analysis helps to interpret the differences and similarities of the data; in fact each PC describes a specific aspect of the variability of the catchment attributes and the variables with the highest loadings on a PC best explain that "dimension" of the data (Chiang et al., 2002; Sanborn and Bledsoe, 2006). The first PC is positively associated with $H_{\min}$, H_{med}, μ_P, σ_P, thus representing the influence of elevation, corresponding to higher rainfall values (due to orographic effect), whereas the correlation scaling exponent, α_P (and also the X coordinate, since the eastern part of the study region is less markedly rugged and receives less rainfall) contributes with

HESSD
9, 10805–10828, 2012

the opposite sign, and therefore with lower values (associated to higher correlation) for increasing altitude. This result is consistent with the findings of Molnar and Burlando (2008), where the most elevated areas exhibit lower α_P-values (and therefore a stronger correlation, indicating that the orographic forcing leads to better organised and long-lasting precipitation fields). The second PC, associated to A, L and $\rho_P(1)$, substantially represents the catchment dimension, increasing along with the lag-1 autocorrelation of the spatially averaged rainfall. The third component, associated with negative X and positive Y-values, represents the geographical location; moving along the Apennine ridge from SE to NW, α_P decreases. This shows that the precipitation on the Emilia area is more correlated than that on the Romagna area, the latter being less mountainous and less rainy, due also to the influence of the sea and of the southern currents.

4.2 Discriminant analysis for classification of ungauged catchments

Discriminant analysis is a supervised learning technique that treats a set of observations with one classification variable and one or more quantitative variables (or discriminants) to describe each classified entity. On the basis of such information, the algorithm constructs a classification rule as a function of the quantitative variables that allows to assign any new record to one of the predefined groups. The analysis identifies the combination of the quantitative variables that maximises the ratio between the inter-classes variance and the intra-classes variance (thus maximising the inter-class separability and the intra-class compactness of the data samples in a low-dimensional vector space), finding the one that can most effectively partition the predefined groups (Hand, 1981; Krzanowski, 1988). In the present application, the quantitative variables describing each entity are the first three principal components of the catchment descriptors (presented in the previous subsection) and the classes are the three clusters identified by the SOM network based on the streamflow signatures. The discriminant capacity is assessed through the comparison between the streamflow signatures classification and the one derived by the catchment attributes. It was here performed

HESSD
9, 10805–10828, 2012

a leave-one-out cross-validation, considering, in turn, each basin as ungauged and therefore excluding it from the data used to construct the discriminant criterion. It is finally possible to determine the percentage of gauged sites correctly classified in the discriminant-based approach. The classification obtained with the discriminant analysis is represented in Fig. 2.

Comparing Figs. 1 and 2, it is highlighted that 9 catchments are misclassified (i.e. an error rate around 20 %): five errors occurred when class 3 entities were assigned to class 2 or viceversa; the remaining four errors result from exchanges between class 1 and class 2. It is worthy observing that there are no instances in which catchments belonging to class 1 are assigned to class 3 or the other way round. This is a merit of the topological properties of the SOM network, unique among the other clustering techniques: the relative position of the nodes on the output layer allows in fact to keep into consideration the affinity among the classes, since nodes that are nearby may be considered representative of akin classes. classes 1 and 3 therefore correspond to the less similar groups, while class 2 is intermediate between the two. It is hence comforting to observe that the two more differing clusters result separate in both classifications and that the only errors are exchanges with the intermediate cluster.

5 Conclusions

The methodology developed in this study first provides a means for identifying groups of similar catchments on the basis of streamflow indexes (signatures) and successively classifies, in the same clusters, ungauged basins on the basis of climate and landscape characteristics. The main novelty of the approach lies in the inclusion, both in the streamflow and in the rainfall characterisation, of information derived by the fine-scale continuous time-series, through indexes attempting to synthesise, in a parsimonious way, the variability and correlation structure of the respective processes.

The streamflow signatures are fed to an unsupervised self-organising mapping network, obtaining three groups with a relatively clear separation, as illustrated also by the

HESSD
9, 10805–10828, 2012

spatial distribution of the members of the classes. In fact, even if no geographical nor elevation information is provided in input to the SOM network, the clusters seem able to distinguish, at least approximately, among parts of the study region coherent as far as location and altitude are concerned and reflecting different climatic and landscape characteristics. Per contra, the values of the indexes chosen for representing the dynamic component of the streamflow process show a not negligible intra-class variability. This was not unexpected because the chosen signatures represent a very simplified description of the autocorrelation function and in particular the assumption of stationarity is indeed a relevant approximation, given the strong seasonality of the streamflow process. The limited number of gauged catchment prevented from applying, in this first study, a more refined representation of the correlation structure. In fact, when limited observations are available, models with too many parameters may result in difficulties for classification and multi-variate analyses as those here presented. It is intended, in future work, to try to enlarge the sample size of available streamflow time-series, for providing a more robust representation of the streamflow process and in particular for testing the potential advantages of a seasonal interpretation of the data.

In order to classify new observations (ungauged sites) to an appropriate streamflow response group, a set of morphologic and pluviometric attributes are identified for describing the catchment. The analysis of the principal components (PCs) of the morpho-pluviometric attributes shows that each of the first three PCs seems able to represent a specific aspect of the differences among the catchments, highlighting, in particular, the role and the dependence among the variables characterising the precipitation regime and its correlation structure.

The results of the discriminant analysis, that identifies the membership of ungauged catchments (described by the corresponding three first PCs) in a leave-one-out cross-validation scheme, evidence a quite satisfactory agreement with the predetermined clusters based on streamflow measures. Moreover, the discriminant analysis clustering is able to clearly distinguish the two less similar groups (class 1 and class 3) identified

HESSD
9, 10805–10828, 2012

in the streamflow-based SOM classification: in fact all the misclassification errors are exchanges with the intermediate class 2.

Of course it is arduous aspiring at a fully appropriate hydrological classification with the set of catchment attributes that are available for this study. For a better characterisation of the phenomena governing the streamflow process, a more comprehensive data set would be needed, including information on the geo-pedological, vegetation and land-use properties of the drainage areas, as well as additional climatic indexes.

Notwithstanding all the above cited limitations, the results confirm the potential of the proposed approach for characterising the catchments. The inclusion of information on the properties of the fine time-scale streamflow and rainfall time-series appears a promising way for better delineating the hydrologic and climatic character of the catchments, at least as far as the present study area is concerned.

Acknowledgements. Analyses are based on measurements provided by the Regional Civil Protection Agency and by the Hydro-Meteorological Service of the Environmental Agency of Emilia-Romagna Region.

References

Atkinson, S. E., Woods, R. A., and Sivapalan, M.: Climate, soil, vegetation controls on water balance model complexity over changing timescales, Water Resour. Res., 38, 1314, doi:10.1029/2002WR001487, 2002. 10807

Bhaskar, N. R. and O'Connor, C. A.: Comparison of method of residuals and cluster analysis for flood regionalization, J. Water Res. Pl.-ASCE, 115, 793–808, 1989. 10816

Burn, D. H.: Cluster analysis as applied to regional floood frequency, J. Water Res. Pl.-ASCE, 115, 567–582, 1989. 10808

Burn, D. H. and Goel, N. K.: The formation of groups for regional flood frequency analysis, J. Hydrol. Sci., 45, 97–112, 2000. 10808

Burn, D. H., Zrinji, Z., and Kowalchuk, M.: Regionalization of catchments for regional flood frequency analysis, J. Hydrol. Eng., 2, 76–82, 1997. 10808

Castellarin, A., Brath, A., and Burn, D.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, J. Hydrol., 241, 270–285, 2001. 10808
Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A.: Calibration of rainfall-runoff models in ungauged basins: a regional maximum likelihood approach, Adv. Water Resour., 2010, 33, 1235–1242, 2010. 10809
Chiang, S. M., Tsay, T. K., and Nix, S. J.: Hydrologic regionalization of watershed, Part I, Methodology development, J. Water Res. Pl.-ASCE, 128, 1–11, 2002. 10808, 10816
Corduas, M.: Clustering streamflow time series for regional classification, J. Hydrol, 407, 73–80, 2011. 10808
De Thomasis, E. and Grimaldi, S.: Introduzione di una metrica tra modelli parametrici lineari nelle applicazioni di tipo idrologico, Giornata di Studio: Metodi Statistici and Matematici per l'Analisi delle Serie Idrologiche, Roma, 2001 (in Italian). 10808
Di Prinzio, M., Castellarin, A., and Toth, E.: Data-driven catchment classification: application to the PUB problem, Hydrol. Earth Syst. Sci. Discuss., 8, 391–427, doi:10.5194/hessd-8-391-2011, 2011. 10812
Haines, A. T., Finlayson, B. L., and McMahon, T. A.: A global classification of river regimes, Appl. Geogr., 8, 255–272, 1988. 10808
Hall, M. J. and Minns, A. W.: The classification of hydrological homogeneous regions, J. Hydrol. Sci., 44, 693–704, 1999. 10812
Hall, M. J., Minns, A. W., and Ashrafuzzaman, A. K. M.: The application of data mining techniques for the regionalisation of hydrological variables, Hydrol. Earth Syst. Sci., 6, 685–694, doi:10.5194/hess-6-685-2002, 2002. 10812
Hand, D. J.: Discrimination and Classification, John Wiley and Sons, Inc., New York, 218 pp., 1981. 10817
Hand, D. J.: Construction and Assessment of Classification Rules, Wiley, Chichester, England, 214 pp., 1997. 10816
Herbst, M. and Casper, M. C.: Towards model evaluation and identification using Self-Organizing Maps, Hydrol. Earth Syst. Sci., 12, 657–667, doi:10.5194/hess-12-657-2008, 2008. 10813
Holmes, M. G. R., Young, A. R., Gustard, A., and Grew, R.: A new approach to estimating mean flow in the UK, Hydrol. Earth Syst. Sci., 6, 709–720, doi:10.5194/hess-6-709-2002, 2002. 10808

HESSD
9, 10805–10828, 2012

Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: An appraisal of the regional flood frequency procedure in the UK "Flood Studies Report", Hydrolog. Sci. J., 30, 85–109, 1985. 10808
Jingyi, Z. and Hall, M. J.: Regional flood frequency analysis for the Gan-Ming River basin in China, J. Hydrol., 296, 98–117, 2004. 10812
Kohonen, T.: Self-Organizing Maps, 2nd Edn., Springer, 1997. 10813
Krzanowski, W. J.: Principles of Multivariate Analysis, A User's perspective, Oxford Univ. Press, New York, 612 pp., 1988. 10817
Laaha, G. and Bloeschl, G.: A comparison of low flow regionalization methods – catchment grouping, J. Hydrol., 323, 193–214, 2006. 10808
Lettenmeier, D. P., Wallis, J. R., and Wood, E. F.: Effect of regional heterogeneity on flood frequency estimation, Water Resour. Res., 23, 313–323, 1987. 10808
Ley, R., Casper, M. C., Hellebrand, H., and Merz, R.: Catchment classification by runoff behaviour with self-organizing maps (SOM), Hydrol. Earth Syst. Sci., 15, 2947–2962, doi:10.5194/hess-15-2947-2011, 2011. 10808, 10812
Lombardi, L., Toth, E., Castellarin, A., Montanari, A., and Brath, A.: Calibration of a rainfall-runoff model at regional scale by optimising river discharge statistics: performance analysis for the average/low flow regime, Phys. Chem. Earth, 42–44, 77–84, 2012. 10809
Marani, M.: On the correlation structure of continuous and discrete point rainfall, Water Resour. Res., 39, 1128, doi:10.1029/2002WR001456, 2003. 10809
McDonnell, J. J. and Woods, R. A.: On the need for catchment classification, J. Hydrol., 299, 2–3, 2004. 10807
Menabde, M., Harris, D., Seed, A., Austin, G., and Stow, D.: Multiscaling properties of rainfall and bounded random cascades, Water Resour. Res., 33, 2823–2830, 1997. 10809, 10811
Merz, R. and Bloeschl, G.: Flood frequency regionalization: spatial proximity vs. catchment attributes, J. Hydrol., 302, 283–306, 2005. 10808, 10815
Molnar, P. and Burlando, P.: Variability in the scale properties of high-resolution precipitation data in the Alpine climate of Switzerland, Water Resour. Res., 44, W10404, doi:10.1029/2007WR006142, 2008. 10809, 10811, 10817
Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: an opportunity for scarcely gauged basins?, Water Resour. Res., 43, W05434, doi:10.1029/2006WR005184, 2007. 10809
Nathan, R. J. and McMahon, T. A.: Identification of homogeneous regions for the purpose of regionalization, J. Hydrol., 121, 217–238,1990. 10808

HESSD
9, 10805–10828, 2012

Catchment classification

E. Toth

Oudin, L., Kay, A., Andreassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, Water Resour. Res., 46, W11558, doi:10.1029/2009WR008887, 2010. 10807

Ozger, M., Mishra, A. K., and Signh, V. P.: Seasonal and spatial variations in the scaling and correlation structure of streamflow data, Hydrol. Process., online first: doi:10.1002/hyp.9314, 2012. 10809, 10811

Patil, S. and Stieglitz, M.: Hydrologic similarity among catchments under variable flow conditions, Hydrol. Earth Syst. Sci., 15, 989–997, doi:10.5194/hess-15-989-2011, 2011. 10808, 10811

Patil, S. and Stieglitz, M.: Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment, Hydrol. Earth Syst. Sci., 16, 551–562, doi:10.5194/hess-16-551-2012, 2012. 10815

Piacentini, T., Castaldini, D., Coratza, P., Farabollini, P., and Miccadei, E.: Geotourism: some examples in Northern-Central Italy, GeoJournal of Tourism and Geosites, 8, 240–262, 2011. 10810

Samaniego, L., Bardossy, A., and Kumar, R.: Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, Water Resour. Res., 46, W02506, doi:10.1029/2008WR007695, 2010. 10807

Sanborn, S. C. and Bledsoe, B. P.: Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, Oregon, J. Hydrol., 325, 241–261, 2006. 10816

Sauquet, E. and Catalogne, C.: Comparison of catchment grouping methods for flow duration curve estimation at ungauged sites in France, Hydrol. Earth Syst. Sci., 15, 2421–2435, doi:10.5194/hess-15-2421-2011, 2011. 10808

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011. 10807

Singh, R. D., Mishra, S. K., and Chowdhary, H.: Regional flowduration models for large number of ungauged Himalayan catchments for planning microhydro projects, J. Hydrol. Eng., 6, 310–316, 2001. 10808

Srinivas, V. V., Tripathi, S., Rao, A. R. and Govindaraju, R. S.: Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering, J. Hydrol., 348, 148–166, 2008. 10812

HESSD

9, 10805–10828, 2012

Catchment classification

E. Toth

Toth, E.: Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting, Hydrol. Earth Syst. Sci., 13, 1555–1566, doi:10.5194/hess-13-1555-2009, 2009. 10813

Vezza, P., Comoglio, C., Rosso, M., and Viglione, A.: Low Flows Regionalization in North-Western Italy, Water Resour. Manag., 24, 4049–4074, 2010. 10808

Viglione, A., Claps, P., and Laio, F.: Mean annual runoff estimation in North-Western Italy, in: Water Resources Assessment and Management Under Water Scarcity Scenarios, edited by: La Loggia, G., CDSU Publ., Milano, Italy, 2006.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment classification and hydrologic similarity, Geogr. Compass, 1, 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007. 10807

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007WR006716, 2008. 10807

HESSD

9, 10805–10828, 2012

Catchment classification

E. Toth

Table 1. Streamflow signatures, pluviometric and morphometric attributes.

Streamflow signatures		Observations Range
Average runoff	μ_Q (mm h^{-1})	0.013–0.233
Standard deviation runoff	σ_Q (mm h^{-1})	0.036–0.588
Percentile 95 % runoff	$P_{Q,95}$ (mm h^{-1})	0.035–0.767
Percentile 5 % runoff	$P_{Q,5}$ (mm h^{-1})	0.000–0.021
Lag-1 autocorrelation runoff	$\rho_Q(1)$	0.962–0.998
Correlation scaling exponent runoff	α_Q	0.088–0.474
Pluviometric attributes		**Observations Range**
Average precipitation	μ_P (mm h^{-1})	0.084–0.258
Standard deviation precipitation	σ_P (mm h^{-1})	0.480–1.309
Proportion of wet hours	P_{Wet}	0.065–0.122
Lag-1 autocorrelation precipitation	$\rho_P(1)$	0.517–0.826
Correlation scaling exponent precipitation	α_P	0.715–1.071
Morphometric attributes		**Observations Range**
Coordinate X UTM streamgauge	X (m)	525 736–758 845
Coordinate Y UTM streamgauge	Y (m)	4 869 659–4 982 633
Drainage area	A (km^2)	18–1303
Minimum catchment elevation	H_{min} (m a.s.l.)	8–896
Average catchment elevation	H_{med} (m a.s.l.)	308–1411
Main stream length	L (km)	3–93

HESSD
9, 10805–10828, 2012

Table 2. Morpho-pluviometric variables with the highest loadings on the first three PCs and total explained variance (in parentheses).

PC_1 (47.8 %)	PC_2 (25.1 %)	PC_3 (16.0 %)
μ_P	A	Y
H_{min}	$\rho_P(1)$	$-X$
σ_P	L	$-\alpha_P$
H_{med}		
$-\alpha_P$, $-X$		

HESSD
9, 10805–10828, 2012

Catchment classification

E. Toth

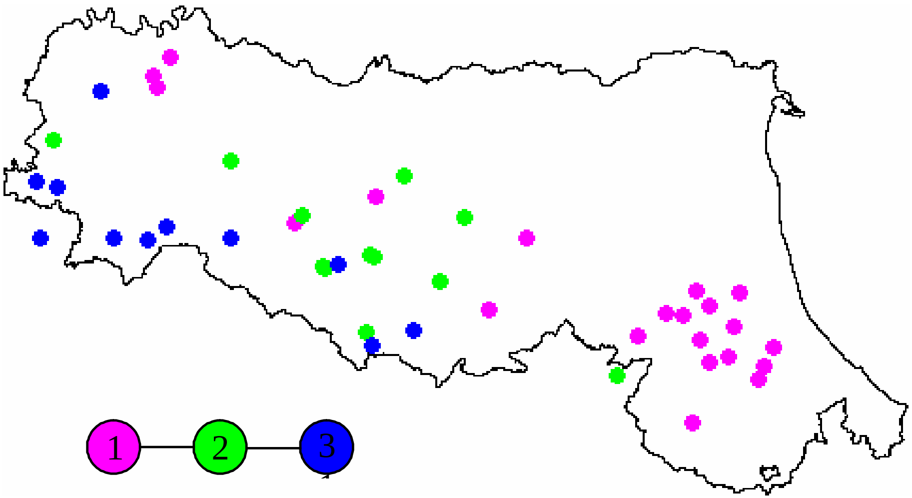

Fig. 1. Classification identified by SOM based on streamflow signatures.

HESSD
9, 10805–10828, 2012

Catchment classification

E. Toth

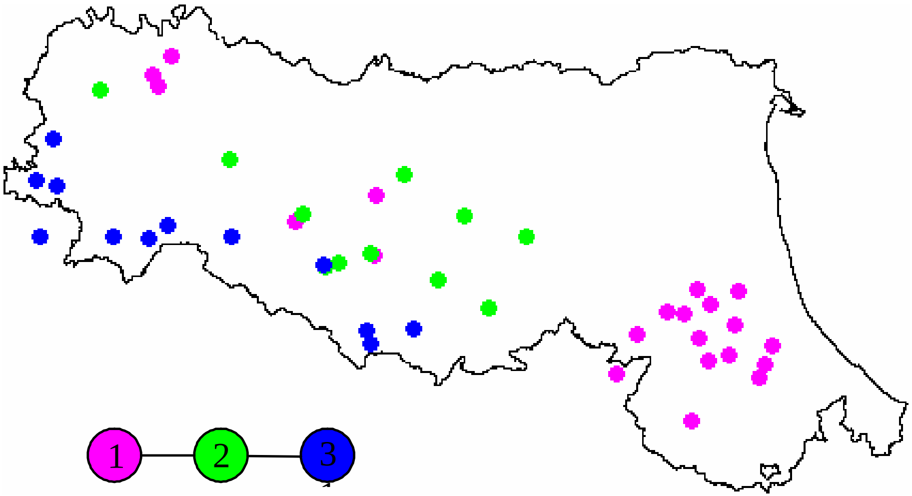

Fig. 2. Classification identified by discriminant analysis based on the first three PCs of the catchment descriptors.

HESSD
9, 10805–10828, 2012

Catchment classification

E. Toth