

## The Influence of Soil Characteristics in Low Flows Regionalization

<sup>1</sup>Paolo Vezza, <sup>2</sup>Claudio Comoglio, <sup>3</sup>Alberto Viglione and <sup>1</sup>Maurizio Rosso

<sup>1</sup>DITIC-Department of Hydraulics, Transport and Civil Infrastructures, Politecnico di Torino,  
Corso Duca Degli Abruzzi, 24, 10129 Torino, Italy

<sup>2</sup>DITAG-Department Land, Environment and Geo-Engineering, Politecnico di Torino,  
Corso Duca Degli Abruzzi, 24, 10129 Torino, Italy

<sup>3</sup>Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology,  
Karlsplatz 13/222 A-1040 Wien, Austria

---

**Abstract: Problem statement:** For the purpose of low flows regionalization, relevant issue for water resources management like environmental flows requirements definition, this study focused on the controls on the seasonal and spatial variability of  $q_{95}$  (i.e., the specific discharge that was exceeded on 95% of all the time) with particular reference to the role of soil characteristics, that, like soil infiltration rate, aquifers recharge, evapotranspiration and topography, usually play a relevant role in low flows seasonality and occurrence within a river. **Approach:** Piemonte and Valle d'Aosta Regions (North-Western Italy) were the investigated study area (30,027 km<sup>2</sup>) where 41 catchments were analyzed with the aim of robust regression models enabling the transfer of hydrological information from gauged to un-gauged sites. **Results:** The regionalization method consisted of multiple regression models between low flows and catchment characteristics. Twenty-five catchment descriptors were used, checking their relative influence with the multi-regressive procedure and a special attention was devoted to the selection of significant soil characteristics in the regionalization process. Seasonality indices were used to classify catchments into two sub-regions and separate multiple regressions was performed by checking the prediction performance with cross-validation. Also a global regression was fitted out but it yielded a lower performance. In the study domain land use, topography and Thornthwaite moisture index demonstrated to be the most significant variables in order to represent relationships between catchment soil characteristics and low flows regime. **Conclusion/Recommendations:** Results obtained in this study were comparable with other regionalization studies carried out in Austria and Switzerland. The interpretation of the identified regression models provided, at local scale, new tools for water management and environmental flows requirements and, from a wider point of view, useful insights into the general comprehension of low flows processes.

**Key words:** Low flows regionalization, catchment descriptors, seasonality indices

---

### INTRODUCTION

An accurate estimate of low flows is needed in many branches of water resources management, including environmental flows requirements. The literature has many references to techniques for estimating low flows regimes at un-gauged sites. In particular catchments are classified in sub-regions according to physiographic characteristics and transferring flow data between catchments in the same region. The most widely used method is the regional regression approach<sup>[9]</sup>. For each sub-region this method relates specific low flows indices (e.g.,  $q_{95}$  that is the

specific discharge that is exceeded on 95% of all the time) to catchment characteristics through linear or non-linear relationships.

Smakhtin<sup>[9]</sup> and Demuth and Young<sup>[2]</sup> give extensive lists of references for regionalizing low flows. In most of these methods, the study domain is divided into sub-regions in which the low flows behavior is assumed to be homogeneous. The identification of these regions is performed by grouping the gauged sites according to a classification criterion and checking the performance of prediction of low flows with cross-validation. In Piemonte and Valle d'Aosta regions (North-Western Italy) low flows regime is characterized

---

**Corresponding Author:** Claudio Comoglio, DITAG-Department of Land, Environment and Geo-Engineering,  
Politecnico di Torino, C. So Duca Degli Abruzzi, 24 10129 Torino, Italy  
Tel: +39.011.5647642 Fax: +39.011.5647686

by two different dry seasons (winter and summer low flows). The consequence is a slow depletion of the soil reservoir in accordance with the recession of discharge within the river. In Alpine climate, winter low flows are affected by freezing processes of snow cover or glaciers. Summer low flows, instead, occur during dry periods when evaporation exceeds precipitation and are normally derived from aquifers recharge. Since recent studies have shown its excellence in classification, seasonality is potentially useful for identifying homogenous groups of catchments<sup>[3,5,7]</sup>. Because of the fundamental differences of summer and winter processes, regionalization may take advantage of a separation into sub-regions based on low flows seasonality.

Gain and losses of stream flow during seasonal dry periods are generated from different natural factors which include soil types and morphological watershed characteristics. Topography, aquifers hydraulics, soil infiltration rate, vegetation distribution and types, geology and climate are basically the most important factors in low flows processes. The driving forces operate on a catchment scale and affect the released discharge from watershed storage<sup>[9]</sup>. The relevance of different 'gain' and 'loss' processes to the wide soil types, topography and climatic conditions, which exist naturally, is far from obvious to determine.

Using a comprehensive multi-regressive approach (i.e., regional regression) it is possible to select the most influential descriptors for the geographic context of the study domain<sup>[11]</sup>. Results of this procedure can ensure complete control of the quality of estimates obtained with regression models based on different kinds of descriptors. In this study, several morphologic and climatic attributes of catchments were selected and computed for 41 river basins in North-Western Italy in order to have an accurate regional estimation of low flows indices.

## MATERIALS AND METHODS

This study was carried out in Piemonte and Valle d'Aosta regions, which have different orographic and climatic characteristics. In this relatively small area (30,027 km<sup>2</sup>) the climate changes from the Apenninic-mediterranean one, with summer low flows in the south-eastern hills, to Alpine-continental one in the Alps mountain range, characterized by winter low flows regime. In this territory catchments areas between 21 and 1,800 km<sup>2</sup>, elevations range from 106-4,725 m and mean annual precipitation from 841 mm in South-Eastern hills to a maximum of 2,113 mm Northern mountainous areas, were considered. For this reason, in

this territory an analysis of soil characteristics influence in low flows regime is both complex and interesting. Figure 1 shows the spatial distribution of stream gauges considered in this study, while Table 1 lists, for each of them, the related catchment area, the Mean Annual Runoff (MAR) and the specific discharge  $q_{95}$ .

**Discharge data:** For discharge data, the reliability of stream flow data series was taken into account: Dams presence in the upstream part of the catchment, a minimum of 10 years of daily stream low registration and relevance of abstractions or karts effects during the low flows periods were the most important factors used in the gauge choice. For instance, gauges located on the plains were excluded owing to flow alterations due to the presence of irrigation abstractions and reservoirs in the upper stream part. Coming from 41 stream gauges, daily discharge data series between 1942 and 1975 were used in this study. As a reference of low flows regime the low flows index  $q_{95}$  (i.e.,  $\Pr(q > q_{95}) = 0.95$ , the discharge exceeded on 95% of all days of the measurement period),

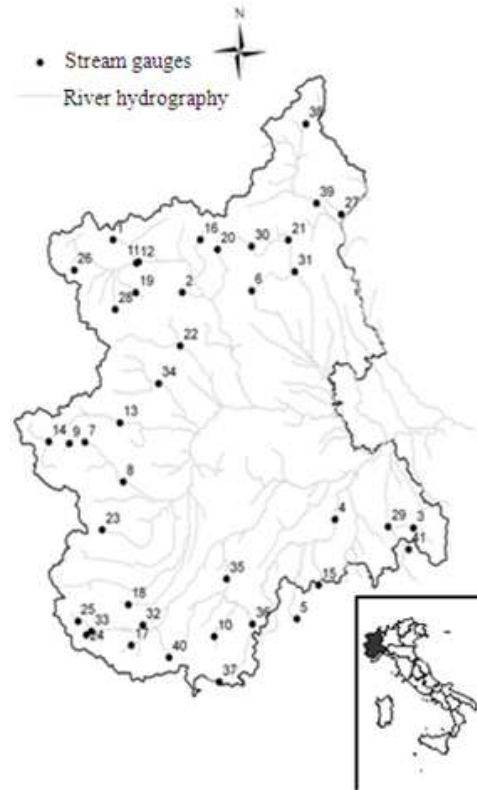


Fig. 1: River network and selected stream gauges in North-Western Italy. As characteristic unit runoff, MAR (mean annual runoff) and  $q_{95}$  are expressed in  $L \text{ sec}^{-1} \text{ km}^{-2}$

Table 1: Stream gauges

Code	Gauge name	Area (km <sup>2</sup> )	MAR (L sec <sup>-1</sup> km <sup>-2</sup> )	q <sub>95</sub> (L sec <sup>-1</sup> km <sup>-2</sup> )
1	Artanavaz St.Oyen	70.90	32.44	10.65
2	Ayasse Champorcher	40.60	39.89	3.43
3	Borbera Baracche	201.80	24.70	2.65
4	Bormida Cassine	1,523.80	16.17	0.98
5	Bormida Mallare Ferrania	49.88	30.60	1.39
6	Cervo Passobreve	75.00	46.33	6.72
7	Chisone Fenestrelle	156.75	20.74	4.53
8	Chisone SMartino	581.00	22.01	5.15
9	Chisone Soucheres Basses	91.94	25.97	5.51
10	Corsaglia Molline	88.94	33.87	3.90
11	Dora Baltea Aosta	1,823.80	28.48	6.33
12	Dora Rhemes Pelaud	54.10	46.07	10.98
13	Dora Riparia Oulx	254.31	21.02	6.71
14	Dora Riparia S.Antonino	992.56	18.74	9.01
15	Erro Sassello	82.69	27.97	2.48
16	Evancon Champoluc	104.90	30.98	5.10
17	GessoValletta S.Lorenzo	110.44	43.89	10.35
18	Grana Monterosso	103.25	25.72	5.55
19	Grand'Eyvia Crétaz	178.60	35.17	3.87
20	Lys Gressoney	90.50	43.03	6.35
21	Mastallone Ponte Folle	146.88	50.74	4.95
22	Orco Ponte Canavese	614.50	32.79	9.66
23	Po Crissolo	37.50	39.76	11.89
24	Rio Bagni Bagni Vinadio	61.63	39.35	9.05
25	Rio Piz Pietraporzio	21.44	40.33	10.04
26	Rutor Promise	45.60	52.26	5.67
27	S. Bernardino Santino	118.81	54.86	8.87
28	Savara Eau Rousee	83.90	34.21	1.08
29	Scrivia Serravalle	616.13	26.22	1.15
30	Sesia Campertogno	169.88	40.43	3.99
31	Sesia Ponte Aranco	702.88	45.28	5.33
32	Stura Demonte Gaiola	560.06	32.06	7.83
33	Stura Demonte Pianche	179.94	29.33	9.36
34	Stura Lanzo Lanzo	576.94	34.56	7.85
35	Tanaro Farigliano	1,516.06	24.61	4.81
36	Tanaro Nucetto	375.63	28.60	4.07
37	Tanaro Ponte Nava	147.63	32.66	3.48
38	Toce Cadarese	189.69	49.82	16.56
39	Toce Candoglia	1,539.81	43.82	15.34
40	Vermenagna Limone	57.44	35.77	8.44
41	Vobbia Vobbietta	56.88	26.48	1.46

was selected because of its large use in literature and its relevance for multiple topics of water resources management<sup>[9]</sup>. q<sub>95</sub> is standardized by the catchment area, (L sec<sup>-1</sup> km<sup>-2</sup>) in order to make it more comparable across scales, expressing characteristic unit runoff. All selected catchments cover a total area of more than 12,000 km<sup>2</sup>, which is about the 40% of the entire study domain.

**Catchment characteristics:** The choice of the catchment descriptors used in low flows regionalization depends largely on the availability and quality of the data. Demuth and Young<sup>[2]</sup> analyzed the frequency of different categories of catchment descriptors in 120 low flows estimation models and assessed that the 73% of all catchment characteristics used are drainage basin parameters, 22% are climatic parameters and 5% are

hydrological parameters. For the drainage basin parameters, 46% are morphometric descriptors, 17% surface cover, 10% geology and 10% soil characteristics.

In this study, for low flows regionalization, 25 morphoclimatic watershed characteristics were used. These descriptors give synthetic information of the shape of basin surface, the nature of the soil and vegetation, the topography and climate (Table 2). Due to the low spatial accuracy of digital information, geology parameters were not considered and land use and runoff curve number were used as geological surrogates<sup>[4,3]</sup>. Also the Thornthwaite moisture index, runoff curve number and the drainage density were included in the analysis because of their relationship with geology, soil infiltration rate and vegetation type distribution. Morphometric parameters of drainage basins and river networks were computed using GIS tools and 'R' statistical software was used for computation of statistical indices<sup>[11]</sup>. Some of the catchment characteristics had to be adapted to make them more useful for regionalization. For instance, it is possible to condense the original Corine land cover classification into 5 land use classes. Drainage basin descriptors were divided into different categories explained using a capitol letter: catchment area A, elevation H, physiographic slope S, orientation O, watershed parameters W, land use L, climatic parameters C. Table 2 shows a summary of these catchment characteristics. Among these, the following are the less common catchment descriptors<sup>[11]</sup>.

Starting from physiographic slope S a mean slope invariant from DEM resolution (S<sub>inv</sub>) was considered, defined as:

$$S_{inv} = \arctg\left(\frac{2(H_{med} - H_{min})}{\sqrt{A}}\right)$$

Where:

A = The catchment area

H<sub>med</sub> = The median elevation

H<sub>min</sub> = The elevation of the closing section

This is a slope measure of a square equivalent basin which does not account for basin shape and whose definition is objective.

For catchment shape factors W<sub>SF</sub> and W<sub>CR</sub> were chosen. The first one expresses the ratio between the catchment area and the square longest drainage path length, while W<sub>CR</sub> is the circularity ratio between drainage basin area and the area of a circle having the same perimeter.

Table 2: Catchment descriptors included in the regional regression analysis. Units were chosen in order to give similar ranges for all characteristics

Symbol	Units	Description	Min	Mean	Max
A	10 <sup>1</sup> km <sup>2</sup>	Catchment area	2.14	35.35	182.38
A <sub>2000</sub>	Percentage	Catchment area above 2000 m	0.00	44.03	97.80
H <sub>max</sub>	10 <sup>2</sup> m	Maximum elevation	9.99	30.56	47.25
H <sub>min</sub>	10 <sup>2</sup> m	Minimum elevation	1.06	8.34	18.80
H <sub>med</sub>	10 <sup>2</sup> m	Mean elevation	4.81	17.76	27.43
S	Percentage	Mean slope	20.20	45.28	63.00
S <sub>inv</sub>	Percentage	Mean invariant slope	0.80	15.53	38.70
S <sub>LDP</sub>	Percentage	Mean slope of the longest drainage path	5.80	18.36	29.30
O	deg	Main catchment orientation angle	1.01	126.15	355.84
O <sub>NORD</sub>	-	Northing (cosine of O)	-0.81	0.43	1.00
O <sub>EST</sub>	-	Easting (sine of O)	-1.00	0.20	1.00
W <sub>DD</sub>	km km <sup>-2</sup>	Stream network density	0.52	0.59	0.74
W <sub>RL</sub>	10 <sup>2</sup> m	Length of the main river	7.17	31.93	133.85
W <sub>LCS</sub>	10 <sup>2</sup> m	Mean length of watershed sides	60.69	74.72	87.06
W <sub>SF</sub>	-	Watershed shape factor	0.08	0.30	0.65
W <sub>CR</sub>	-	Watershed circularity ratio	0.24	0.49	0.74
L <sub>U</sub>	Percentage	Urbanized areas	0.00	0.11	0.75
L <sub>F</sub>	Percentage	Forested areas	0.47	46.94	99.96
L <sub>CG</sub>	Percentage	Crop and grassland	0.04	11.08	53.21
L <sub>R</sub>	Percentage	Wasteland (rocks)	0.00	41.83	99.35
L <sub>W</sub>	Percentage	Wetland	0.00	0.04	1.17
L <sub>CN</sub>	-	Runoff curve number	26.32	42.19	50.06
C <sub>P</sub>	10 <sup>2</sup> mm	Mean annual precipitation	8.41	12.62	21.13
C <sub>IT</sub>	-	Thornthwaite moisture index	0.04	0.89	1.92
C <sub>IB</sub>	-	Budyko aridity index	0.45	0.85	1.20

In the land use section L, the runoff curve number (L<sub>CN</sub>) is an empirical parameter used in hydrology for predicting direct runoff or infiltration from rainfall excess. L<sub>CN</sub> is based on the hydrological soil groups, land use, treatment and hydrological moisture condition. The curve number method was developed by the USDA Natural Resources Conservation Service, which was formerly called the Soil Conservation Service (SCS). In literature L<sub>CN</sub> is also known as ‘SCS runoff curve number’<sup>[10]</sup>.

For climatic parameters C, the Thornthwaite index (C<sub>IT</sub>) and the Budyko index (C<sub>IB</sub>) were considered. C<sub>IT</sub> is a global moisture index that can be estimated as the ratio:

$$C_{IT} = \frac{C_p - ET_0}{C_p}$$

Where:

ET<sub>0</sub> = The mean annual potential evapotranspiration on the basin

C<sub>IB</sub> = Instead, is a radiational aridity index expressed as:

$$C_{IB} = \frac{R_n}{\lambda \cdot C_p}$$

Where:

R<sub>n</sub> = The mean annual net radiation

λ = The latent vaporization heat

Values assumed by C<sub>IB</sub> are lower than 1 for humid regions and greater than 1 in arid regions<sup>[11]</sup>.

**Regional regression analysis:** The regional regression analysis was performed in two step: The first is to divide the study domain into sub-regions in which the low flows behavior is assumed to be homogeneous, while the second is to build a multi-regressive model that relates the q<sub>95</sub> (i.e., dependent variable) to morphoclimatic descriptors (i.e., independent variables) in order to select the most influential descriptors for low flows regionalization.

Recent studies have shown the reliability of seasonality indices method for classifying catchments in order to divide the study domain. Laaha and Blöschl<sup>[6]</sup> investigated four catchment grouping strategies when developing multi-regressive models to estimate low flows indices in Austria. Due to the large difference in low flows occurrence a catchment grouping based on seasonality gave the best performance. Engeland and Hisdal<sup>[3]</sup> compared a regional regression model with a regional rainfall-runoff model in Norway for regionalizing low flows. In the regional regression they used seasonality indices to divide the territory into sub-regions. Also in that case regression method generally gives better estimates of low flows in un-gauged catchments.

**Seasonality indices:** In order to investigate the low flows seasonality three indices, as in Laaha and Blöschl<sup>[7]</sup>, were used.

The first one is the Seasonality Ratio (SR) which expresses the difference of summer ( $q_{95s}$ ) and winter ( $q_{95w}$ ) low flows. SR is defined as:

$$SR = \frac{q_{95s}}{q_{95w}}$$

From April 1st to November 30th daily discharge time-series are considered as winter discharges and from December 1st to March 31st as summer discharges. Values of  $SR > 1$  indicate the presence of a winter low flows regime and values of  $SR < 1$  indicate the presence of summer low flows regime (Fig. 2a).

The second seasonality parameter is composed by two indices  $\theta$  and  $r$ <sup>[7,12]</sup>. These represent the seasonal distribution of low flows occurrence. The parameter  $\theta$  is a circular statistic. Its values range between 0 and  $2\pi$ ,

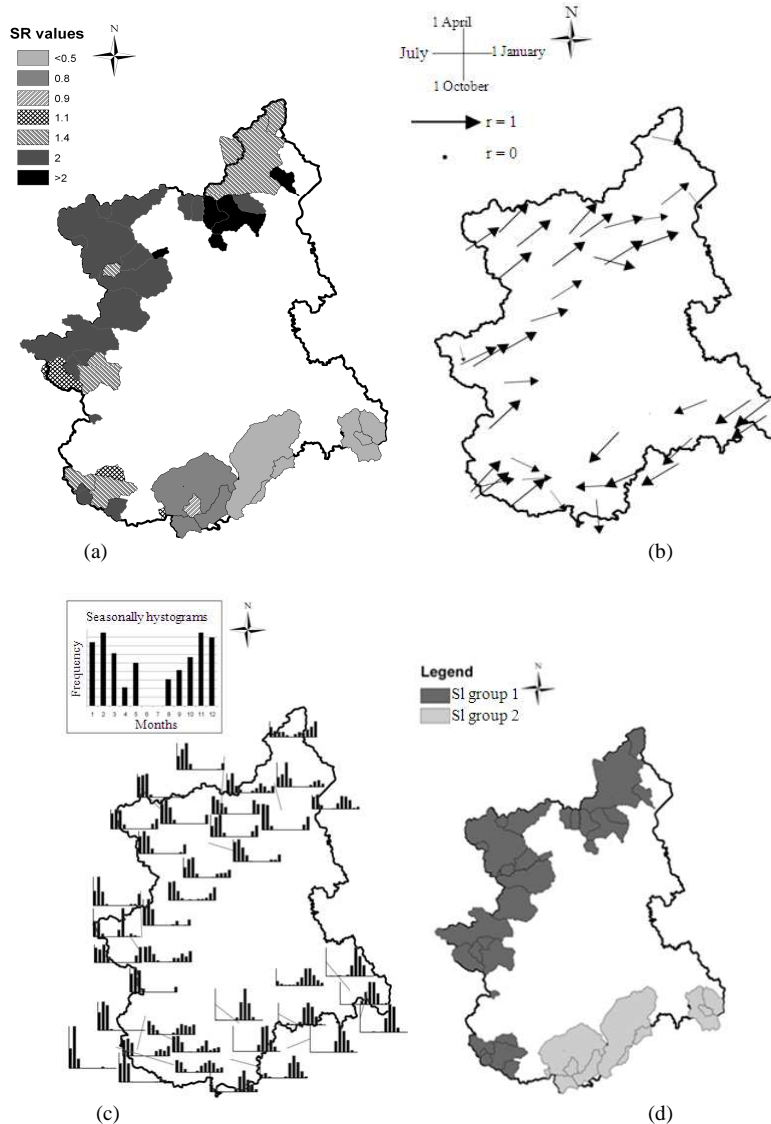


Fig. 2: Seasonality indices analysis: (a): Seasonality Ratio (SR) between summer and winter low flows discharges; (b): Variability and mean day of occurrence of  $q_{95}$ ; (c): Non-exceedance frequency histograms (SHs) of specific low flow  $q_{95}$ ; (d): Groups of catchments based on seasonality indices

explaining the  $q_{95}$  mean day of occurrence (e.g.,  $\theta = 0$  relates to January 1st,  $\pi/2$  relates to April 1st,  $\pi$  relates to July 1st and  $3\pi/2$  relates to October 1st). The parameter  $r$  shows the variability of low flows seasonality. It ranges from zero to unity, where  $r = 1$  corresponds to strong seasonality (all low flows events occurred on exactly the same day of the year) and  $r = 0$  corresponds to no seasonality (low flows events are uniformly distributed over the year). Indices are calculated from Julian dates  $D_j$  (values on the range 1-365 or 1-366 in leap years) of all days of the observation period when discharges are equal or below  $q_{95}$ .  $\theta_j$  is the  $D_j$  directional angle expressed in radians as:

$$\theta_j = \frac{D_j 2\pi}{365}$$

The mean direction angle expressing the  $q_{95}$  mean day of occurrence is defined by:

$$\theta = \arctan\left(\frac{y_0}{x_0}\right) \text{ 1st and 4th quadrants: } x > 0$$

$$\theta = \arctan\left(\frac{y_0}{x_0}\right) + \pi \text{ 2nd and 3rd quadrants: } x < 0$$

where,  $x_0$  and  $y_0$  are the mean Cartesian coordinates for a total of  $n$  single days  $j$ :

$$x_0 = \frac{1}{n} \sum_j \cos(\theta_j)$$

$$y_0 = \frac{1}{n} \sum_j \sin(\theta_j)$$

In the end, the measure of the variability of low flows seasonality is expressed by the length of the mean vector  $r$ :

$$r = \sqrt{x_0^2 + y_0^2}$$

With  $\theta$  and  $r$  the mean day of occurrence and the intensity of seasonality can be displayed by using a vector map (Fig. 2b).

The third seasonality index is expressed with seasonality histograms based on a monthly scale. Columns represent the frequency of discharges below the threshold  $q_{95}$  over time (Fig. 2c). Regions with approximately homogeneous seasonality are shown in

Fig. 2d. Group 1 includes the Alpine region, characterized by winter low flows and Group 2 is referred to the Apenninic-mediterranean climate where low flows normally occur during summer.

**Low flows indices estimation:** Discharge data used in this study were daily discharges. To make the low flows characteristics more comparable across scales,  $q_{95}$  was standardized by the catchment area ( $q_{95}$  [ $l \text{ s}^{-1} \text{ km}^{-2}$ ]). The types of linear models investigated were divided in four classes:

$$q_{95} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} + \varepsilon$$

$$q_{95}^{1/2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} + \varepsilon$$

$$q_{95}^{1/3} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} + \varepsilon$$

$$\ln(q_{95}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} + \varepsilon$$

Where:

$x_i$  = The morphoclimatic descriptors

$\beta_i$  = Regression coefficients

The analysis of low flows data series showed that distribution of  $q_{95}$  and its transformations resulted in approximate normality. For the estimation of the coefficients  $\beta_i$  the Ordinary Least Squares technique<sup>[8]</sup> was used. For all regression models, a combination of all morphoclimatic variables was attempted, satisfying the three general assumptions: The absence of multicollinearity between  $\beta_i$ , the homoscedasticity ( $\text{Var}[\text{res}_i] = \text{const.}$ ) and the unbiasedness ( $E[\text{res}_i] = 0$ , guaranteed by the OLS procedure) of residuals, where  $\text{res}_i$  is the residual for catchment  $i$ . All the models for which at least one of the independent variables resulted to be non-significant according to the Student  $t$  test at a 95% significance level were discarded. The descriptive power of each regression was assessed through the adjusted coefficient of determination  $R^2_{\text{adj}}$ , defined as:

$$R^2_{\text{adj}} = 1 - \frac{(n-1) \sum_{i=1}^n (q_{95,i} - \hat{q}_{95,i})^2}{(n-p) \sum_{i=1}^n (q_{95,i} - \bar{q}_{95,i})^2}$$

Where:

$n$  = The number of considered stations

$p$  = The number of estimated coefficients

$q_{95}$  and  $\hat{q}_{95,i}$  = The measured and estimated mean annual flow at the  $i$ -th site

$\bar{q}_{95,i}$  = The average of the mean annual flows for all considered gauges

The coefficient of determination  $R^2_{adj}$  is useful to choose the best model among the ones belonging to a given class but cannot be used to compare models of different nature. To this purpose a cross-validation method was carried out, computing the RMSE (Root Mean Square Error) on the residuals  $\hat{q}_{95,i} - q_{95,i}$ , where  $\hat{q}_{95,i}$  is the estimated value of the  $i$ -th dependent variable obtained using a model estimated with all the observations except the  $i$ -th one. The  $RMSE_{CV}$  is defined as the square-root of the average residual square error  $V_{CV}$ :

$$RMSE_{CV} = \sqrt{V_{CV}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{q}_{95,i} - q_{95,i})^2}$$

Finally, it is possible to define the coefficient of determination based on cross-validation as:

$$R^2_{CV} = \frac{Var(q_{95}) - V_{CV}}{Var(q_{95})}$$

For each class the multi-regressive model based on the best performances in terms of  $R^2_{adj}$ , with the lower  $RMSE_{CV}$  (the best model) and with the use of the most commonly-available parameters (the simplest model) was chosen. Finally the selected model was then checked with respect to the assumptions underlying the regression analysis.

Multi-collinearity affects the OLS procedure determining large variances and co-variances for the least-squares estimators of the regression coefficients. A simple statistic to measure the presence of multi-collinearity is the Variance Inflation Factor (e.g.<sup>[8]</sup>):

$$VIF = (1 - R_j^2)^{-1}$$

Where:

$R_j^2$  = The coefficient of determination obtained when the independent variable

$x_i$  = Regressed on the remaining  $p-1$  regressors

Practical experience indicates that if any of the VIFs exceeds 5 or 10, the associated regression coefficients are poorly estimated because of multi-collinearity.

In the end normality of residuals is required for hypothesis testing (the significance  $t$  test). To detect non-normality and heteroscedasticity of the residuals, they were plotted respectively on a normal probability

plot and against the fitted values, in order to recognize if they display particular patterns.

As a first approach, one global regression model was fitted to all 41 catchments. In the second step, corresponding to the original classification of catchments obtained by seasonality indices analysis, regionally restricted regression models were each fitted for the two contiguous regions (Fig. 2d).

## RESULTS

Best regressions were chosen on the basis of the criteria previously discussed considering all the possible linear regression models. Table 3 outlines the best regressions obtained for each model classes, along with  $R^2$  and  $RMSE_{CV}$  statistics. Chosen multi-regressive relationships and their performance are reported in Table 4, which also outlines two measures of model performance, the coefficient of determination  $R^2_{CV}$  and the root mean squared error  $RMSE_{CV}$ . Both were obtained from cross-validated residuals and, therefore, are representative of the prediction of low flows in ungauged catchments.

Table 3: Best regression obtained between the  $q_{95}$  (dependent variable) and morphoclimatic parameters (independent variables) with  $R^2$  and  $RMSE_{CV}$  statistics

Dependent variable	Model	Independent	$R^2_{adj}$ variable	$RMSE_{CV}$
$q_{95}$	Global	$A_{2000} S_{LDP} L_{CG} C_{IT}$	0.669	2.357
	Gr. 1	$A_{2000} W_{RL} L_{CG} C_{IT}$	0.647	2.207
	Gr. 2	$H_{max} W_{RL} L_F C_{IT}$	0.960	1.612
$q_{95}^{(1/2)}$	Global	$A_{2000} L_{CG} L_{CN}$	0.669	2.378
	Gr. 1	$A_{2000} W_{RL} L_{CG}$	0.442	2.319
	Gr. 2	$H_{med} S_{LDP} L_{CG}$	0.893	1.826
$q_{95}^{(1/3)}$	Global	$A_{2000} S L_{CG}$	0.680	2.336
	Gr. 1	$A_{2000} S L_{CG}$	0.568	2.640
	Gr. 2	$H_{med} S_{LDP}$	0.818	1.950
$\ln(q_{95})$	Global	$A A_{2000} L_{CG} L_{CN}$	0.639	2.986
	Gr. 1	$A_{2000} S W_{RL} L_{CG}$	0.561	2.807
	Gr. 2	$H_{med} S_{LDP}$	0.826	1.775

Table 4: Regression coefficients and performance of selected regional regression models for  $q_{95}$  ( $L \text{ sec}^{-1} \text{ km}^{-2}$ ). For the units of catchment characteristics (Table 1)

Model	$q_{95}$ regression model	$R^2_{adj}$	$RMSE_{CV}$	$RMSE_{CV-TOT}$	$R^2_{CV}$
Global	$\sqrt[3]{q_{95}} = 0.8130 + 0.0037 \cdot A_{2000} + 0.0146 \cdot S + 0.0132 \cdot L_{CG}$	0.68	2.336	2.336	0.60
Gr. 1	$q_{95} = -6.6656 + 0.0780 \cdot A_{2000} + 0.09137 \cdot W_{RL} + 0.1916 \cdot L_{CG} + 4.5929 \cdot C_{IT}$	0.65	2.207	1.931	0.72
	$q_{95} = 5.0551 + 0.2478 \cdot H_{max} - 0.020653 \cdot W_{RL} + 0.0463 \cdot L_F - 2.5648 \cdot C_{IT}$	0.96	1.612		

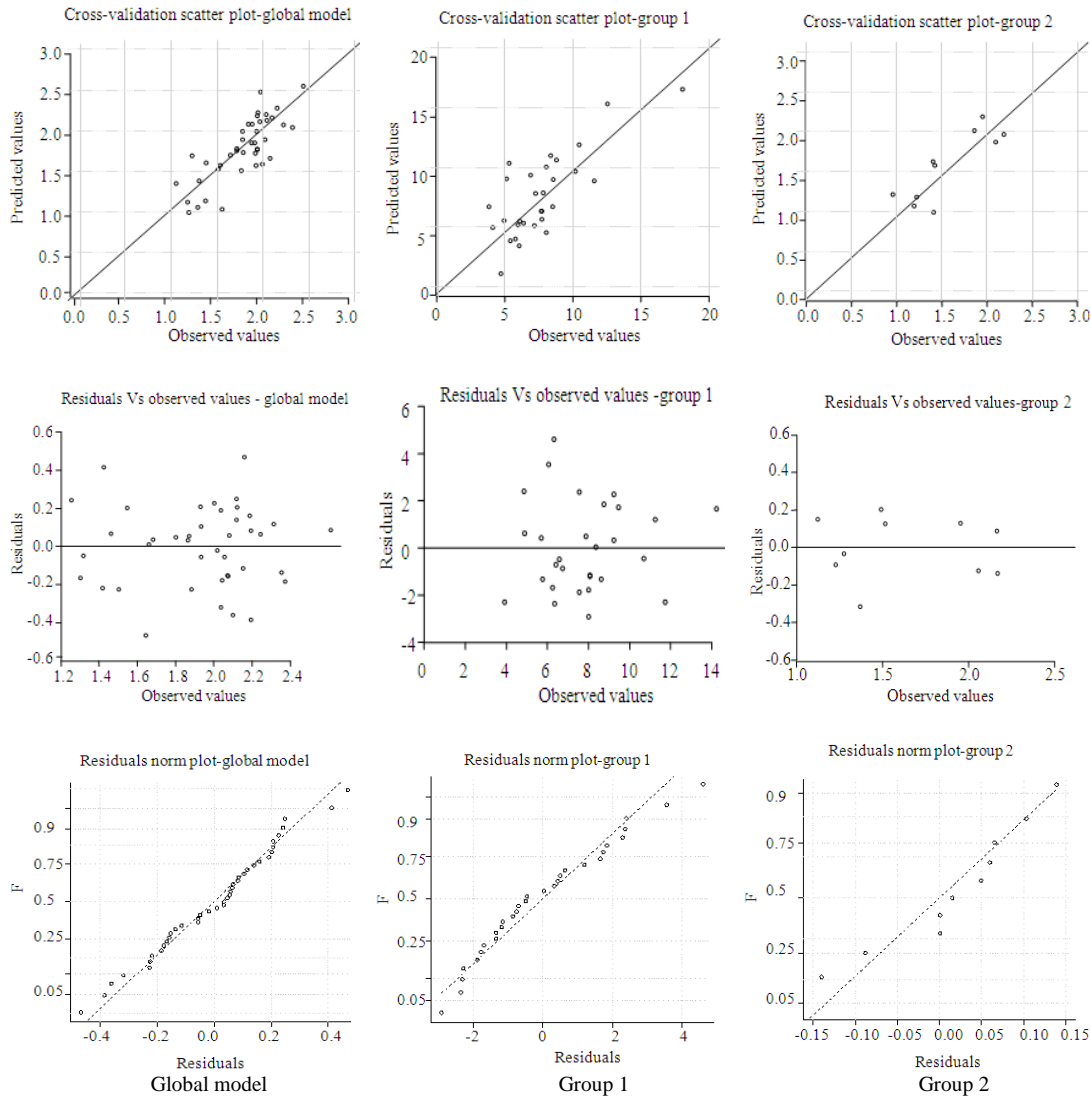


Fig. 3: Diagnostic plots of residuals. Each column refers to one regression model (global and sub-regional models)

Global regression showed a relative performance of  $R^2_{cv} = 60\%$ , corresponding to  $RMSE_{CV} = 2.336 \text{ L sec}^{-1} \text{ km}^{-2}$ . Grouping catchments into two sub-regions and separate regressions using seasonality indices improved the overall model performance to  $R^2_{cv} = 72\%$  and  $RMSE_{CV} = 1.1931 \text{ L sec}^{-1} \text{ km}^{-2}$ .

Normality and homoscedasticity of residuals are desirable properties if one is interested in interpretable estimates of model performance. In this study, model assumptions (i.e., normality of residuals and heteroscedasticity) were carefully checked by three diagnostic graphs: Scatter plots of observed versus predicted values, residual plot as a function of observed

values and normal probability plots of residuals. Diagnostic graph derived by the regression results are showed in Fig. 3. Each column corresponds to one regional regression model and each point to one catchment. Proceeding by rows, the first two graphs allow a detailed examination of the performance of individual catchments, including the existence of outliers and a potential heteroscedasticity (diversity in variance) of the residuals. In the graphs the outliers do not tend to increase with  $q_{95}$  and the residuals of the three models can be considered homoscedastic. In the context of this study, results of this check are positive for all regression models, as the main focus is on



evaluating the influence of soil characteristics on low flows regionalization.

The third graph is the representation of cross-validated residual in normal probability plot. Only single extreme outliers appear and residuals can be considered approximately normally distributed. In order to check for multi-collinearity, the VIF factor was computed for all the regressors considered. Factor values ranges between 1.02 and 2.65 for all descriptors considered in regression models. In all cases VIF is much below 5, value indicating possible multi-collinearity.

**Relative importance of catchments descriptors:** A more detailed analysis of Table 3 and 4 yields the relative importance of predictor variables in the context of low flows regionalization. Starting from the global regression model, parameters importance consists of three catchment characteristics. Percentage of catchment area above 2000 m ( $A_{2000}$ ), mean slope ( $S$ ) and the proportion of crops and grassland ( $L_{CG}$ ) demonstrate to be the most significant variables for the transfer of hydrological information. Especially for the geographic context under exam,  $L_{CG}$  (i.e., proportion of crop-grasslands) appears in all selected regression models and it is the most influential descriptor for land use.

Grouping into two regions and separate regressions in each region led to two similar regression equations in terms of considered descriptors. Models exhibit quite the same parameters. Land use, river length and Thornthwaite index appear in both models as the most influential descriptors. The two different sub-regional models differ because of changes in selected regressors. Particularly, in the regression models of Group 2,  $A_{2000}$  and  $L_{CG}$  are replaced by similar descriptors  $H_{max}$  (maximum elevation) and  $L_F$  (proportion of forest).

In the context of this study very little influence on low flows appears to exist for soil characteristics such as runoff curve number, stream network density and portion of urbanized areas, rocks and wetlands. They are never selected in the regression equations.

## DISCUSSION

Soil characteristics and especially land use demonstrate to be significant variables for the regionalization of low flows indices. It is interesting to compare this result with studies in the literature using similar catchments characteristics and examined  $q_{95}$  specific discharges.

The frequency of the catchment descriptors used depends largely on the availability and quality of the

data. In literature catchment descriptors affecting low flows controls are comprehensively provided by Smakhtin<sup>[9]</sup>. In that study it is possible to find an overview of catchment descriptors used in regional estimation models. Catchment area, mean annual precipitation, channel and/or catchment slope, stream frequency and/or density, percentage of lakes and forested areas, various soil and geology indices, length of the main stream, catchment shape, watershed perimeter and mean catchment elevation are the morphoclimatic characteristics most commonly used.

In this study 25 catchment descriptors were used, checking their relative influence with the multi-regressive procedure. In the resulted regional regression models, only 7 catchment characteristics were selected. Moreover, as shown in Table 2, regression models contained similar parameters and most of them occurred for different regression classes.

The overall regression model yielded a lowest performance ( $R^2_{cv} = 60\%$ ). For this global model three catchment characteristics were selected as predictors:  $A_{2000}$  (catchment area above 2000 m),  $S$  (mean catchment slope) and  $L_{CG}$  (proportion of crop-grasslands). In Alpine climate  $A_{2000}$  is related to the catchment topography and consequently to the special influence in snow cover and glaciers formation. These processes characterize the whole Alps mountain range in Northern Italy and regulate the hydrological response during winter low flows periods. Smakhtin<sup>[9]</sup> give an extensive list of low flows studies related to the influence of glaciers and snow pack. Freezing and melting processes include a decrease in runoff variation and, consequently, more sustained low flows. The mean slope  $S$  generally has a positive effect on low flows<sup>[7]</sup> and it is possibly correlated with storage volume in high mountains.

Grouping catchments into two sub-regions had the best performance ( $R^2_{cv} = 72\%$ ). Separate regressions using seasonality indices in catchment classification were carried out. Catchment relief was represented also in separate grouping equations by  $A_{2000}$  or maximum altitude ( $H_{max}$ ). Hence, altitude and slope (catchment topography) appears as an important control of low flows in the study domain. Other catchment characteristics that appeared into sub-regional regressions are  $W_{RL}$  (river length),  $L_F$  (proportion of forest) and  $C_{IT}$  (Thornthwaite moisture index). River length relates to the longest drainage path and its interactions with aquifers. Engeland and Hissdal<sup>[3]</sup> find main river length as a reference for winter low flows regime in Norway.

Land use is significant for all selected models in the study domain. It affects evapotranspiration and soil infiltration rate which are related to water losses in low flows discharges during dry periods. The proportion of forest, crops and grasslands ( $L_F$  and  $L_{CG}$  respectively) are the land-use characteristics which more frequently appear in regressions. In contrast with Laaha and Blöschl<sup>[7]</sup> portion of rocks and wetland were never selected and did not appear to be important controls. On the other hand, Aschwanden and Kan<sup>[1]</sup> within a study concerning the low flows regionalization for Switzerland, outlined that land use plays an important role in predicting low flows indices, especially considering the characteristics proportion of agricultural areas and pre-Alpine farming structures.

The climatic parameter  $C_{IT}$  (the Thornthwaite index) expresses the ratio between mean annual precipitation and evapotranspiration. It relates to the catchment global moisture as a reference of the average water balance. It was selected in the two separate regression models, supporting the finding that Thornthwaite index is one of the most important controls of low flows in North-Western Italy.

Catchment area, orientation parameters, watershed shape factors, runoff curve number and Budyko aridity index were never selected in the regression equation, so very little influence on low flows appears to exist. According to Laaha and Blöschl<sup>[7]</sup>, also stream density  $W_{DD}$  (parameter related to geology and land use) never occurred in the regression equations and does not appear to be a significant indicator of low flows for the study domain.

Results obtained in this study are therefore comparable with regionalization studies in literature, especially in Austria and Switzerland. We believe that the interpretation of the regression models provides useful insights into the comprehension of low flows processes. Topography, Thornthwaite moisture index and land use conditions appear to be the most influential parameters for regionalizing low flows in North-Western Italy.

## CONCLUSION

Gain and losses of stream-flow during seasonal low flows periods are generated from different natural factors that include morphoclimatic descriptors of catchments. The objective of this study was the examination of the relative influence of soil characteristics in regionalizing low flows. In this study, the specific discharge  $q_{95}$  is considered as an index of low flows regime with regard to its estimation in un-

gauged basins within the Piemonte and Valle d'Aosta regions (North-Western Italy). Regionalization has a primary importance in water resources management in order to transfer hydrological information from gauged to un-gauged sites. Multiple regression analysis were used and tested in order to establish a relationship between catchment characteristics and the low flows index  $q_{95}$ . Based on the results, using regression method, catchment grouping based on the dominant low flows seasonality was an effective method for obtaining homogeneous sub-regions in North-Western Italy where winter low flows and summer low flows are controlled by different processes. The best results in this study were obtained with an overall predicting performance of  $R^2_{cv} = 72\%$ . Topography, Thornthwaite moisture index and land use conditions are the best index for determining low flows estimates for ungauged catchments. According to other studies in literature, it can be affirmed that these kinds of analyses are useful in order to assess the relative importance of different kind of descriptors. Moreover, the interpretation of soil and land use parameters using regression models provides useful insights into the comprehension of low flows processes.

## ACKNOWLEDGEMENT

This research has been funded by the Regione Piemonte, C61 Project CIPE 2004. The Authors wish to acknowledge W. Mattalia and G. Bertolotto, Regione Piemonte officers, for their precious support.

## REFERENCES

1. Aschwanden, H. and C. Kan, 1999. The discharge Q347-State of the question. Communications hydrologiques 27. Swiss National Hydrological and Geological Survey, Berne. <http://www.bafu.admin.ch/publikationen/publikation/00397/index.html?lang=fr>
2. Demuth, S. and A.H. Young, 2004. Regionalization Procedures. In: Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater, Tallaksen, L.M. and H.A.J. Van Lanen (Eds.). Developments in Water Science 48, Elsevier, ISBN: 0-444-51688-3, pp: 307-343.
3. Engeland, K. and H. Hirdal, 2009. A comparison of low flow estimates in ungauged catchments using regional regression and HBV-Model. Water Res. Manage., 1: 1-20. DOI: 10.1007/s11269-008-9397-7

4. Gustard, A. and R. Gross, 1989. Low Flow Regimes of Northern and Western Europe. In: FRIENDS in Hydrology, Gustard, A. and R. Gross (Eds.). IAHS Press, Wallingford, ISBN: 0-947-57127-2, pp: 516.
5. Laaha, G., 2002. Modelling Summer and Winter Droughts as a basis for Estimating River Low Flows. Proceeding of the international Conference on Regional Hydrology: Bridging the Gap between Research and Practice, Mar. 18-22, IAHS Press, Wallingford, pp: 289-295. <http://cat.inist.fr/?aModele=afficheN&cpsid=14055919>
6. Laaha, G. and G. Blöschl, 2005. Low flow estimates from short stream flow records: A comparison of methods. *J. Hydrol.*, 306: 264-286. <http://cat.inist.fr/?aModele=afficheN&cpsid=16787368>
7. Laaha, G. and G. Blöschl, 2006. Seasonality indices for regionalizing low flows. *Hydrol. Proc.*, 20: 3851-3878. <http://cat.inist.fr/?aModele=afficheN&cpsid=18316933>
8. Montgomery, D., E. Peck and G. Vining, 2001. Introduction to Linear Regression Analysis. 3rd Edn., Wiley, New York, ISBN: 10: 0471315656, pp: 672.
9. Smakhtin, V.U., 2001. Low flow hydrology: A Review. *J. Hydrol.*, 240: 147-186. <http://researchspace.csir.co.za/dspace/handle/10204/549>
10. United States Department of Agriculture, 1986. Urban hydrology for small watersheds. Technical Release 55 (TR-55). <https://people.sunysit.edu/~barans/classes/ctc260/pp/55/TR55.ppt>
11. Viglione A., P. Claps and F. Laio, 2007. Mean Annual Runoff Estimation in North-Western Italy. In: Water Resources Assessment Under Water Scarcity Scenarios, La Loggia G., G. Aronica and G. Ciruolo (Eds.). CSDU Italy, ISBN: 978-88-900282-8-1, pp: 97-122.
12. Young, A.R., C.E. Round and A. Gustard, 2000. Spatial and temporal variations in the occurrence of low flow events in the UK. *Hydrol. Earth Syst. Sci.*, 4: 35-45. <http://hal.archives-ouvertes.fr/docs/00/30/45/01/PDF/hess-4-35-2000.pdf>