Regional analysis for rivers low flow statistics estimation

Rossi, Giuseppe
University of Firenze, Department of Civil and Environmental Engineering
Via S. Marta, 3
50139 Firenze, ITALY
E-mail: giuseppe.rossi@dicea.unifi.it

Caporali, Enrica
University of Firenze, Department of Civil and Environmental Engineering
Via S. Marta, 3
50139 Firenze, ITALY
E-mail: enrica.caporali@unifi.it

Introduction

Low flows are normally derived from groundwater discharge or surface discharge from lakes, marshes, or melting glaciers. Lowest annual flow usually occurs in the same season each year. The natural factors which influence the various aspects of the low-flow regime of the river include the distribution and infiltration characteristics of the soils, the hydraulic characteristics and extent of the aquifers, the rate, frequency and amount of recharge, the evapotranspiration rates from the basin, the distribution of vegetation types, topography and climate (Smakhtin, 2001).

Low flow characteristics are estimated from observed streamflow data, identifying duration curves, indices and percentiles characteristics. Although various low-flow indices describe different aspects of low-flow regime of a river, most of them are obviously strongly intercorrelated. Two main groups of low flow indices are usually used in drought identification. In the first group are the percentiles (Q70 and Q90) derived from the Flow Duration Curve (FDC). The second group is composed by the minimum n-day average discharge indices (i.e. Q (7, 2) and Q (7, 10)) (Pyrce, 2004).

Low flow regime is closely dependent on the catchment area hydrogeological features but on a practical perspective, although scientifically proven, statistical analysis is widely applied to derive indices to characterize low flow regimes. Indices are commonly evaluated at gauged sites from observed streamflow time series. To improve their reliability, often affected by the lack of observed data and to estimate low flow statistics in ungauged sites it is possible to refer to regional statistical analysis. The method employs catchment area and climatic characteristics, as independent variables, and data from other catchments where stream flow data are recorded. It is the most widely used technique in flow estimation in ungauged sites or where few data are available (Riggs, 1973). The regional analysis improves the capability to predict the water flow regime at gauged sites with short time series, reducing the uncertainties and moreover allows the estimation of the discharge properties at ungauged sites (Chokmani and Ouarda, 2004). In the regional analysis the data from all sites in a region are evaluated to define regions that are hydrologically homogeneous in terms of characteristic being studied. Regional analysis of extremes would require advances in the methodology of the statistics (Katz et al., 2002). The application of frequency analysis for hydrological extremes evaluation has a long history in hydrology.

The proposed methodology for low flows regionalization consists of the following steps: selection and analysis of recorded data; discharge frequency analysis; definition of homogeneous regions; discharge estimation; evaluation of the procedure. The basic procedure has been described in the quoted studies for flood peak discharges and has been modified in this study for the regionalization of low flows.

The analysis of low flow indices is carried out on the discharge data of 65 consistent hydrometric stations located in Tuscany Region. The area is subdivided into different regions using the L-moments method applied to the 7-day annual minima and to the Q70 annual series. The subdivision is tested using discordancy and heterogeneity statistics. For each river section two interpolation techniques, either deterministic (Inverse Weighted Distance) or geostatistical (Ordinary Kriging), are applied. The results are validated using the Jackknife method to assess the Root Mean Square Error RMSE.

Study area and dataset

The analysis is carried out on the discharge data recorded in several rivers in the Tuscany Region central Italy (Figure 1), having an area of about 23˙000 (Regione Toscana, 2008). The main rivers of the region are: Arno, Serchio, and Ombrone Grossetano. Moreover there are small basins of coastal rivers near the Tyrrhenian Sea and the upstream part of Tevere, Fiora and Magra watersheds.

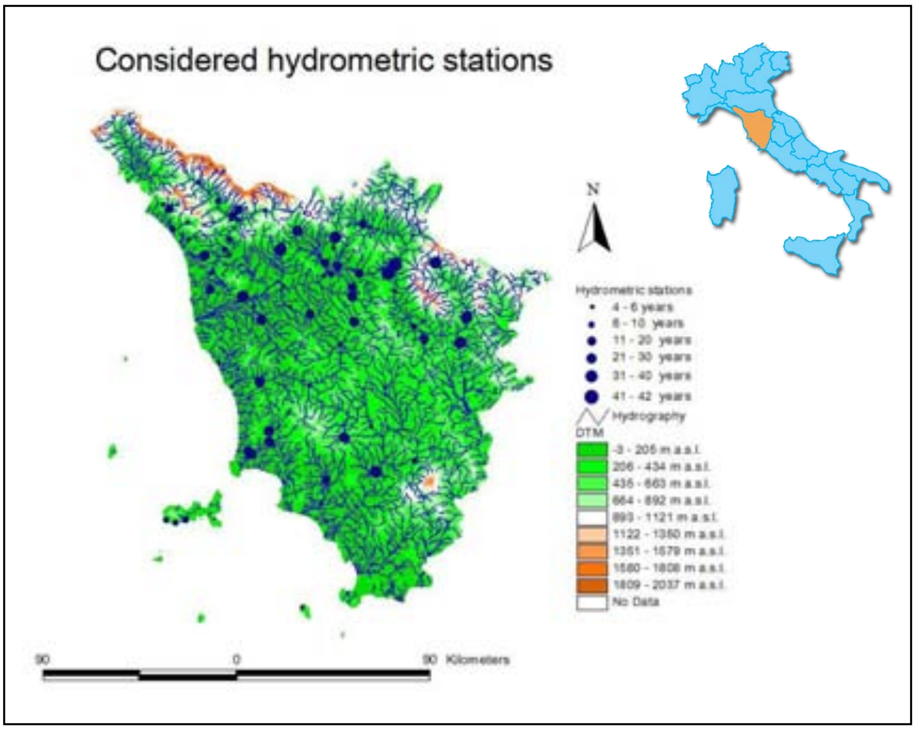

Figure 1. *The Tuscany Region and the considered hydrometric stations with the years of registrations. In blue the hydrographic network.*

The dataset used in the study was registered by Servizio Idrologico Regionale Toscano (Regional Hydrologic Service of Tuscany) using data of the network previously managed by Ufficio Idrografico e Mareografico (Hydrographic and Mareographic Office) integrated with a new network during the period 1949 to 2008. Excluding all the stations with less than 3 years of data, not using stations with long periods of inactivity and merging the data of traditional analogical and digital automatic stations, when they are placed in the same location, a dataset of 65 stations was obtained.

Methodology

The first step in regionalisation studies is the delineation of hydrologically and statistically homogeneous regions. In some cases it is clear how to group a domain into regions of approximately uniform hydrological and statistical behavior but, more often, the choice is far from obvious (Laaha and Bloeschl, 2006). Homogeneous regions can be defined as geographically contiguous regions, geographically non contiguous regions, or as hydrological neighborhoods. To employ geographically contiguous regions is easier than using non contiguous regions, especially in the context of scarcity of data. On the other hand, even two adjacent river catchments may have different topography, soils or other local anomalies (Laaha and Bloeschl, 2005). In this work we decided to use contiguous regions and to delineate them using boundaries based on physiographic considerations. The aim was to find hydrologically and statistically homogeneous regions in the area of interest, using standardized low flow characteristics from available observed stream flow records (1949– 2008) for the Tuscany Region, central Italy. Following this, a low flow event regional frequency analysis, based on L-moments was carried out.

Low flow events are represented here by two indices: the 7-day annual minimum series and the annual Q70 series. After testing and arranging the data, various indices were calculated. In Table 1 are reported some statistical characteristics of flow indices for the 65 considered catchments

Table 1. *Statistical characteristics of low flow indices for the 65 considered catchments.*

Index	**unit**	**minimum**	**median**	**mean**	**maximum**
Q70	**m$^{3/}$s**	0.000	0.543	4.265	118.933
Q90	**m^{3}/s**	0.000	0.185	2.934	102.982
Q(7,2)	**m^{3}/s**	0.000	0.129	2.444	92.138
Q70/A	**l/(s*km^{2})**	0.000	2.241	5.940	62.592
Q90/A	**l/(s*km^{2})**	0.000	0.998	3.196	56.678
Q(7,2)/A	**l/(s*km^{2})**	0.000	0.655	2.062	11.744

The L-moments approach was used to assign these data to the different regions, according to homogeneity measures and properties. This method is based on useful statistics for regional frequency analysis, which measure regional homogeneity and goodness-of-fit and it is proposed by Hosking and Wallis (1993), based on L-moments method defined by Hosking (1990).

The L-moments are expectations of linear combinations of order statistics that were defined by Hosking (1990) as linear combinations of probability weighted moments (PWMs), previously introduced by Greenwood et al. (1979). Four parameters are evaluated to identify those sites that are discordant with the group as a whole. The first one is the Discordancy statistics, measured in terms of the first four L-moments of the sites' data. The other three parameters, the Heterogeneity statistics, are relative respectively to the first, first and second, first and third L-moment (Hosking and Wallis, 1993).

Large values of D_i indicate sites that are most discordant from the group. In this work, a site is considered to be unusual if the Discordancy measure (D_i) is larger than 3 and possibly discordant if D_i is larger than 2. Large values of *Hi* indicate region with sites that are really discordant from the group. A region is homogenous if any of the *Hi* values is less than 1, possibly heterogeneous if *Hi* is between 1 and 2, and definitely heterogeneous if *Hi* is more than 2 (Hosking and Wallis, 1993).

Discussion of results

The described method is applied to the Tuscany region dataset. The L-moments for the two selected low flow indices are calculated and once the area of interest is divided into different regions, geographically contiguous, the homogeneity measures are calculated to test each subdivision. Proceeding by trial-and-error some sub-basins were moved from one region to another, and some regions were split into sub-regions to reach the best possible homogeneity. The discordancy (D_i) and the heterogeneity (*H1, H2,* and *H3*) are calculated firstly for the whole area considered as a unique region. Values of calculated homogeneity statistics suggested that this approximation was not correct. In particular 5 stations have values of Discordancy higher than 3 (Table 2), the threshold value of the discordancy measure and *H1* for the whole region has a value that is considerably higher than 2, the threshold levels to consider a region "definitely heterogeneous". The area is successively split into three different sub-regions, following previous studies on rainfall extreme values (Tartaglia et al., 2006; Caporali et al., 2008). With this subdivision there is some homogeneity, but some stations still present high values of discordancy. Only the North sub-region has a value of *H1* that is above the "definitely heterogeneous" threshold level.

Table 2. *Values of the homogeneity parameters for the 7-day annual minimum series. In red are shown the parameters that define a "definitely heterogeneous region", in bold the ones that define a "possible heterogeneous" region.*

Regions	Number of stations	*H1*	*H2*	*H3*	Number of sites *D*>2	Number of sites *D*>3
Unique	48	**3.89**	0.99	**1.58**	**9**	**5**
North	21	**2.13**	**1.25**	**1.62**	**4**	**2**
Centre	21	**1.58**	0.99	**1.71**	**4**	**2**
South	6	**1.63**	0.70	0.94	**1**	0
North East	11	0.43	0.74	0.34	**1**	0
North West	9	**1.22**	0.79	**1.36**	**1**	**1**
Centre East	11	**1.16**	0.77	0.96	0	0
Centre West	9	**1.80**	0.85	**1.33**	**3**	0
South	7	**1.60**	0.77	0.95	0	0

Finally a new subdivision into 5 sub-regions is proposed (Figure 2), splitting the central and the northern regions of the previous subdivision. Some stations are moved from one sub-region of the previous subdivision to another one. Once the gauge stations belonging to the same sub-region are individuated, the different sub-regions are delimitated following the main hydrological watersheds. The station of Colonna is not included in the subdivision, due to non-homogeneity of its data. With this subdivision the regions are more homogeneous, and the subdivision follows hydrological and precipitation features.

The previous subdivisions are tested even with the annual Q70 values (Table 3). Due to the homogeneity of these values, the subdivision into three regions seems sufficient and, since that only North and Centre sub-regions are above the "possibly heterogeneous" threshold levels and no stations have a D_i value above 3. The subdivision into five sub-regions gives better results.

Table 3. *Values of the homogeneity parameters for the Q70 annual series. In red are shown the parameters that define a "definitely heterogeneous region", in bold the ones that define a "possible heterogeneous" region.*

Regions	Number of stations	*H1*	*H2*	*H3*	Number of sites *D*>2	Number of sites *D*>3
Unique	48	**2.22**	0.66	0.90	**3**	**2**
North	21	**1.43**	0.64	0.87	0	0
Centre	21	**1.03**	0.59	0.89	**1**	0
South	6	0.81	**1.04**	0.97	**1**	0
North East	11	0.27	0.31	0.31	0	0
North West	9	**1.28**	0.52	0.76	0	0
Centre Eas	11	0.60	0.41	0.61	0	0
Centre Wes	9	**1.38**	0.61	0.76	**2**	0
South	7	0.70	0.88	0.84	0	0

An appropriate interpolation technique over the geographical space have to be established. The first considered interpolation technique is the Inverse Distance Weighted (IDW). IDW methods are based on the assumption that the interpolating surface should be influenced most by the nearby points and less by the more distant points. The interpolating surface is a weighted average of the scatter points and the weight diminishes as the distance from the interpolation point to the scatter point increases.

The second considered interpolation technique is a geostatistical method, the Ordinary Kriging. Geostatistical methods are powerful statistical techniques designed to study spatially autocorrelated variables (Isaaks and Srivistava, 1989). They permit estimating the local value of a variable using sparse local measurements. Between the several variograms a spherical variogram is used, as in other flows interpolations proposed in literature (i.e Castiglioni et al., 2008).

The results are validated using the Jackknife method (Tukey, 1958) based on removing data and then recalculating the estimator. The root mean square error is assessed for the three proposed subdivisions and for both the proposed low flow indices to compare the results, and to quantify the accuracy of the different techniques. In Table 4 are reported the RMSE values for the two considered interpolation techniques.

Results confirm the good properties of homogeneity of the final subdivision for three sub-regions (South, Centre East, and Centre West) while for other two (North East and North West) the results are not the expected ones. For the North East region it probably depends on the variability of the values while for the North West region for the different geo-climatic characteristics that are not taken into account with this interpolation. With the Ordinary Kriging interpolations results are really similar to the one found with the IDW even if there is an improving of results especially for the northern regions.

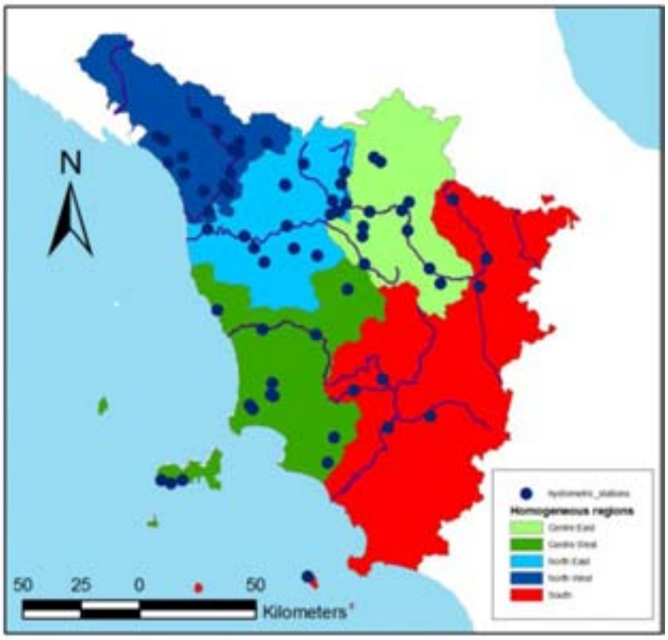

Figure 2 *Subdivision in hydrologically and statistically homogeneous regions.*

Table 4. *Values of the RMSE - Root Mean Square Error for IDW and Ordinary Kriging interpolations.*

Subdivision	Regions	IDW RMSE Q(7,2)	IDW RMSE Q70	Ord. Kriging RMSE Q(7,2)	Ord. Kriging RMSE Q70
Unique	Unique	3.19	9.25	3.02	8.65
3 regions	North	4.02	13.44	3.89	11.33
	Centre	2.61	3.72	2.60	3.42
	South	0.75	2.26	0.74	2.08
5 regions	North East	4.14	15.20	3.94	10.69
	North West	4.10	10.76	3.82	11.06
	Centre East	0.58	0.64	0.58	0.51
	Centre West	0.63	0.86	0.61	1.01
	South	0.70	2.08	0.69	1.84
	Mean	2.76	8.96	2.54	7.23

Conclusions

The Tuscany Region rivers low flows are analysed and a subdivision in homogeneous regions is evaluated with the L-moments method and with hydrologic characteristics of the studied area. Three different subdivisions are tested. A unique region is evaluated, but it is not sufficiently homogeneous. The subdivision into three different sub-regions, following previous studies on rainfall extreme values gives some homogeneity, but some stations still presented high values of discordancy. Finally a new subdivision with 5 sub-regions is proposed following the main hydrological watersheds. This subdivision reaches a good degree of homogeneity. Low flow indices at ungauged basins are evaluated through Inverse Weighted Distance and Universal Kriging interpolations.

The results are validated using the Jackknife method and calculating the RMSE – Root Mean Square Error for the different techniques and the different subdivisions. For IDW the RMSE values confirm the good properties of homogeneity of the final subdivision for three sub-regions (South, Centre East, and Centre West) while for other two (North East and North West) the results are not the expected ones. Ordinary Kriging performs better, especially in the North East and North West sub-regions.

REFERENCES (RÉFERENCES)

Caporali E., Cavigli E., Petrucci A. (2008). The index rainfall in the regional frequency analysis of extreme events in Tuscany (Italy). *Environmetrics*. 19: 714-724.

Castiglioni S., Castellarin A., Montanari A. (2008). Stima delle portate di magra in siti non strumentati mediante tecniche di interpolazione spaziale (Low flow estimation in ungauged sites through spatial interpolation techniques). *Proceedings of 31° Convegno Nazionale di Idraulica e Costruzioni Idrauliche Perugia, 9-12 settembre 2008*. 8 pp.

Chokmani K.., Ouarda T. B. M. J. (2004). Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resources Research.* 40: 1–12.

Greenwood J. A., Landwahr J. M,. Matalas N. C., Wallis J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research.* 15: 1049–1054.

Hosking J. R. M. (1990). L-moments: analyzing and estimation of distributions using linear combinations of order statistics. *Royal Statistical Society Journal, Series B.* 52: 105–124.

Hosking J. R. M., Wallis J. R. (1993). Some statistical useful in regional frequency analysis. *Water Resources Research.* 29: 271–281.

Isaaks E. H., Srivistava R. M. (1989). *Applied Geostatistics*, Oxford University Press. 569 pp.

Katz R. W., Parlange M. B., Naveau P. (2002). Statistics of extremes in hydrology. Advances in Water Resources 25: 1287–1304.

Laaha G., Bloeschl G. (2005). Low flow estimates from short stream flow records-a comparison of methods. *Journal of Hydrology.* 306: 264–286.

Laaha G., Bloeschl G. (2006). A comparison of low flow regionalization methods – catchment grouping. *Journal of Hydrology.* 323: 193–214.

Pyrce R. (2004). *Hydrological low flow indices and their uses*. Watershed Science Centre Report No. 04-2004. Trent University. Canada. pp.38.

Regione Toscana, Unioncamere Toscana, ISTAT – Istituto Nazionale di statistica (2009). *Annuario Statistico Regionale, Toscana 2008 (Regional statistic yearbook, Toscana 2008)*. Pubblicazioni del Sistema Statistico Regionale.

Riggs H. C. (1973). *Regional analysis of streamflow characteristics.* US Geological Survey Techniques of Water Resources, United States Government Printing office, Washington. VIII+15 pp.

Smakhtin V. Y. (2001). Low flow hydrology: a review. *Journal of Hydrology*. 240: 147-186.

Tartaglia V., Caporali E., Cavigli E., Moro A. (2006). L-moments based assessment of a mixture model for frequency analysis of rainfall extremes. *Advances in Geosciences.* 2: 331–334.

Tukey J. W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics.* 29: 614.