Climate regimes and yearly streamflow frequency analysis in Sicily

C.Arena[1] & M.R. Mazzola[1]
[1]Dipartimento di Ingegneria Idraulica ed Applicazioni Ambientali
Università di Palermo
Viale delle Scienze, Edificio 8 – 90128 Palermo
ITALY
E-mail: claudioarena1@yahoo.it

Abstract

Yearly streamflow frequency analysis is performed in Sicily, Mediterranean's largest island, using the index flow procedure, i.e. fitting a single probability distribution function (pdf) to the available yearly flow data scaled by their mean. This requires 1) the identification of homogeneous regions where L-moments' observed variability can be ascribed to sample variability only, and 2) the estimation of mean annual streamflow by multiple regression analysis with some significant climatic and morphologic covariates. Rather than having defined geographical boundaries, the identified homogeneous regions should gather basins sharing similar morphologic and climatic features. Steps 1) and 2) are commonly kept separated; however, classic multiple regression analysis has a potential to identify regions with different average streamflow levels through measures of heteroscedasticity. This can be related to the presence of climatic regimes and is useful to direct analysis in the identification of basin typologies that can be characterized by a single pdf. In Sicily, within a general Mediterranean climate, different climatic regimes co-exist ranging from humid to semiarid. Tests for heteroscedasticity on the residuals of a regression of mean yearly flow versus the morphologic and climatic covariates indicate that groupwise heteroscedasticity exists at a regional scale and therefore suggest splitting the sample in two subsamples of semiarid and subhumid basins. What is significant, however, is that such subdivision, based on Thornthwaite's humidity index, also proves effective from the standpoint of the homogeneity of the pdf of yearly flows. Further analysis is carried out to test the goodness-of-fit of different candidate distributions of the scaled yearly streamflow.

1. INTRODUCTION

Yearly streamflow is a key variable in many engineering and scientific applications, ranging from reservoir design and simulation to soil water balance in order to estimate aquifer recharge. Regional water resources planning often makes use of estimates of streamflow and its variability at many sites, for some of which very scarce or no measures at all may be available. In these cases, regional analysis is a widely used tool for the assessment of mean yearly streamflow and its variability (Vogel *et al*, 1999). The idea is to use multiple regression analysis to infer mean yearly streamflow and variance at an ungauged site using the hydroclimatic and geomorphic information available over an area. When matched with an appropriate probabilistic model of yearly streamflow, i.e., with a probability density function (pdf hereafter), or even better with a stochastic process with a given marginal pdf, such techniques provide the hydrologic input necessary to risk analyses in various fields of water resources engineering.

A different approach is based on the index flood procedure (Darlymple, 1960), a technique that has now found numberless applications worldwide for the frequency analysis of flood events and that easily lends itself to be applied to the frequency analysis of yearly streamflow (Viglione *et al*. 2006). Following Viglione *et al*. (2006), we shall denote such approach as the index flow procedure. Clearly, some difference exists between the two processes, i.e. yearly peak flow and yearly streamflow, as in general terms annual streamflow is both a serially and cross correlated process, whereas the assumption of independence in time certainly holds in the case of yearly peak flows. In many instances, however, serial correlation can be neglected also when analyzing yearly streamflow, as in the case of Sicily where, owing to the fact that most rivers are ephemeral, almost all of the available series display lag one serial coefficients that are statistically non significantly different from zero.

The idea behind the index flow procedure is that, given an homogeneous region where N streamflow records are available, the growth curve, (that is, the pdf of yearly streamflows scaled by an index flow -

the mean annual streamflow, for instance), is the same for all the available records: the estimated L-moments of the streamflow records are different one from the other only because of sampling variability.
The three main steps to be taken to perform the whole procedure can be summarized as follows:

- Identify homogeneous regions;
- Select some candidate pdfs for the growth curve, estimate its parameters and test its goodness of fit;
- Estimate the index flow (the mean annual streamflow) by multiple regression analysis on climatic and geomorphic covariates in order to assess the quantiles of yearly streamflow at sites where no runoff measurements are available.

The first step requires the use of a criterion to identify areas (groups of sites) with the same growth curve. One commonly employed measure of homogeneity is the Hosking and Wallis' H (Hosking & Wallis, 1993) that is obtained by simulation, as will be illustrated in section 3. Here it may suffice to remark that such homogeneous regions need not have defined geographical boundaries, but should rather gather basins with similar morphologic or climatic characteristics so that the homogeneity assumption may have some physical ground. As recommended by Hosking & Wallis (1993), never should homogeneous regions be identified by gathering or clustering sites based on the L-moment values, as this would result in a tautology and add no information to the analysis. In order to identify such regions, tests based on clustering techniques such as the Mantel test (Viglione *et al.,* 2006) have been proposed. However, we will try to show that multiple regression analysis for the assessment of mean yearly flow already contains a lot of information for the identification of regions where the assumption of a single growth curve may hold. To this end, we set the multiple regression problems in a spatial framework to show how simple measures of heteroscedasticity can be used to identify areas with a homogeneous streamflow level and then check such areas for homogeneity from the standpoint of the growth curve through Hosking and Wallis' index H.

2. MEAN ANNUAL STREAMFLOW AS A SPATIAL PROCESS

As anticipated, mean annual streamflow at engaged sites is usually assessed by regressing the long-term mean yearly streamflow at N available measuring sites with the geomorphic and climatic features of the basins. It can be useful to set such regressive analysis in a spatial framework: mean annual streamflow MAF(**s**) at a site $s \in R^2$ can be viewed as a spatial process of the type (Cressie, 1993):

$$\mathrm{MAF}(\mathbf{s}) = \boldsymbol{\mu} + \delta(\mathbf{s}) \quad (1)$$

where $\boldsymbol{\mu}$ and $\delta(\mathbf{s})$ respectively represent the large-scale and the small-scale variation of the process. Large scale variation can be expressed as $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, i.e. as a linear function of q explanatory variables (in this case precipitation, mean elevation, average hillslope etc.). **X** is $n \times q$ matrix, being n the number of realizations of the spatial process (i.e. the number of sites where mean annual streamflow estimates are available) and $\boldsymbol{\beta}$ is the parameter vector. $\delta(\mathbf{s})$ is in general a zero – mean process with a $\boldsymbol{\Sigma}$ semi-positive definite variance – covariance matrix.
If $\delta \sim N(0, \sigma^2\mathbf{I})$, then the process is homoscedastic and if residuals are also uncorrelated in space, $E[\boldsymbol{\varepsilon\varepsilon}^T] = 0$, then the general spatial model reduces to the classic non-spatial i.i.d error model and OLS (Ordinary Least Squares) estimates of $\boldsymbol{\beta}$ can be used. However, there are instances in which residuals at different sites are uncorrelated, yet the assumption $\delta \sim N(0, \sigma^2\mathbf{I})$, does not hold. In such cases, $\boldsymbol{\Sigma}$ is still a diagonal matrix to be written as $\boldsymbol{\Sigma} = \mathbf{I}\boldsymbol{\sigma}$ with $\boldsymbol{\sigma}^T = [\sigma_1, \sigma_2 \ldots \sigma_n]$.
In the econometric literature, where such spatial regression models are popular, this is often referred to as structural instability or structural change and is related to spatial heterogeneity that appears when a variable distributes itself differently through the spatial units; these ideas can be easily transferred to the analysis of mean annual flow heterogeneity among river basins in a given geographical area: if $\boldsymbol{\Sigma} = \mathbf{I}\boldsymbol{\sigma}$, the assumption of a fixed relation between the explanatory variables and the dependent variable that holds over the complete dataset is not tenable. Instead, heterogeneity may be present, in the form of different intercepts and/or slopes in the regression equation for subsets of the data. When the different subsets in the data correspond to regions or spatial clusters, it is

referred to a switching regression specification as spatial regimes. If for instance k different spatial regimes exist over the study area, the model of mean annual streamflow can be written in this fashion:

$$\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \dots \\ \mathbf{D}_k \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \dots & \dots & .. & \dots \\ 0 & 0 & 0 & X_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \dots \\ \boldsymbol{\beta}_k \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \dots \\ \boldsymbol{\varepsilon}_k \end{bmatrix} \quad (2)$$

where vector $\mathbf{D_i}$ = [D_{i1}, D_{i2}….$D_{is(i)}$] contains the mean annual flow in the s(i) sites of the i-th spatial regime, so that $\sum_{i=1}^{k} s(i)$ = n. Matrix **X** contains the explanatory variables of the process, $\boldsymbol{\beta}_\mathbf{i}$ = [β_{i1}, β_{i2}….$\beta_{iq(i)}$] is the vector of model parameters for the i-th regime and $\boldsymbol{\varepsilon}_\mathbf{i}$ = [ε_{i1}, ε_{i2}….$\varepsilon_{is(i)}$]is the vector of residuals of the i-th spatial regime.

Moran's I is a statistic that can be used to detect autocorrelation among residuals of a regression. It is defined as follows:

$$I = \frac{\hat{\boldsymbol{\varepsilon}}^T \cdot \frac{1}{2}\left(\mathbf{W} + \mathbf{W}^T\right) \cdot \hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}^T \cdot \hat{\boldsymbol{\varepsilon}}} \quad (3)$$

Where $\hat{\varepsilon}^T$ is the vector of the residuals of the regression and **W** is a symmetric matrix of standardized spatial weights defining the relationship among pairs of spatial objects. For instance, element w_{ij} of matrix **W** may equal the inverse of the distance between residual at site i and residual at site j.The most common approach to employ I as a test for spatial autocorrelation is to assume that under the null hypothesis (that is: no spatial autocorrelation is present), I is distributed as a normal variable, with E[I] = -1/(N –1) and VAR[I] as a function of sample size, E[I] and matrix **W** (Cliff & Ord, 1981, page 18). If the null assumption cannot be rejected, a diagonal matrix model for **Σ** can be accepted. However, it must be still checked whether a form like $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ is suitable or a form like $\mathbf{\Sigma} = \mathbf{I}\boldsymbol{\sigma}$ should be adopted. Tests on heteroscedasticity can help in this regard.

A number of tests are available in the literature to assess whether residuals of a regression display non-constant variance and hence heteroscedasticity. One commonly (and generic) used model for non constant variance is to assume that the variance of residuals is proportional to the square of one explanatory variable. Based on this assumption, the Goldfeld – Quandt test requires the identification of an explanatory variable as a source for heteroscedasticity. The values of such variable are ranked in ascending order and the corresponding observations are arranged accordingly. Two separate regressions are performed using the first and last third of the data. For each regression the RSS_i (the sum of squared residuals) i = 1, 2, is obtained.The test statistic is GQ = RSS_1/RSS_2 which is distributed as a Fisher's F with (T_1 – q) and (T_2 – q) degrees of freedom. T_1 and T_2 are the size of the two samples used for the regression and q the number of the explanatory variables.

Unlike the Goldfeld – Quandt test, the White test makes no assumption on the variable X_i and works with all the q explanatory variables: possible sources of heteroscedasticity are identified in the explanatory variables, in their squares and in their cross products. The test hence assesses if a significant correlation actually exists among the squared residuals of the regression and the explanatory variables, their squares and their cross products by testing the joint hypothesis that the coefficients of the regression among εi^2 (dependent variable), X_{ij}, Z_{ij}^2 and X_iX_j are all zero. Under this assumption, the statistic W =nR^2 (n is the sample size) is χ^2 distributed with p-1 degrees of freedom, where p is the number of explanatory variables potentially explaining the variance of residuals that is linked to the number q of explanatory variables in the regression by p = q(q + 3)/2.

3. IDENTIFICATION OF HOMOGENEOUS AREAS IN THE INDEX FLOW PROCEDURE

The evaluation of the homogeneity of an area from the standpoint of the marginal pdf of indexed annual streamflow is carried out in the framework introduced some fifteen years ago by Hosking and Wallis (1993): given an homogeneous area for which streamflow samples at N different sites are available, each of different length n_i, i = 1, 2...N and with sample mean $\hat{\mu}_i$, i = 1, 2...N, consider Q_i, the 100%P percentile of the marginal pdf of yearly streamflow at site i. According to the index flow procedure, $Q_i(P) = \hat{\mu}_i$ q(F) where q(F) is a growth curve that holds for the whole homogeneous area. The moments of the q(F) as estimated from the various samples can hence be different only because of the sampling variability. The procedure makes use of the L-moments instead of the classic product moments, as L-moments are more robust to outliers and do not have sample-size related bounds, and can be applied to both the second-order L-moment ratio (the L-coefficient of variation), to the third (asymmetry) and fourth (kurtosis).

The homogeneity of N yearly streamflow samples scaled by their mean can be assessed in terms of L-Cv and L-asimmetry. To this end a parent distribution for the scaled data must be assumed; Hosking & Wallis (1993) suggest that a quite general distribution Q(F) such as the four parameter kappa (Hosking, 1994) be used. The kappa distribution has the following form:

$$Q(F) = \xi + \alpha/k \{1 - [(1 - F^h)/h]^k\} \quad (4)$$

Parameter estimation for such distribution is performed following the Hosking & Wallis (1997) scheme, that is, by estimating the first four PWMs and subsequently the L-moments for each of the N sites. The four parameter of the distribution are linked to its first four moments by relationships available in Hosking (1994) constituting a nonlinear system to be solved by iterative methods.

Once the parameters of the Q(F) have been estimated, a large number of worlds (500 in this application) are simulated from the distribution. Here "world" indicates a simulated set of samples with each set having the same number of samples of the same lenght as the one available in the real world. For each of such simulated worlds, the variance $V_j^{(sim)}$ j = 1,...500 of the sample L-cv is assessed, weighted by the length n_i of the samples:

$$V_j^{(sim)} = \sum_{i=1}^{N} n_i * (Lcv_i - \underline{Lcv_j})^2 \quad (5)$$

Where $\underline{Lcv_j}$ denotes the average L-cv for the j-th world. Denote with μ_V and σ_V the mean and the standard deviation of such 500 $Vj^{(sim)}$. The measure of homogeneity is given by

$$H = \frac{V^{(obs)} - \mu_V}{\sigma_V} \quad (6)$$

Where V(obs) indicates the variance of L-cv calculated by (5) from the real world sample. Intuitively, the smaller the value of H, the likelier the assumption that the scaled streamflow available in the observed samples are drawn from the same parent distribution.

Hosking & Wallis suggest that values of H < 1 should indicate that the selected area is homogeneous, H > 2 indicates heterogeneity while values between 1 and 2 point out the possibility of heterogeneity for the area considered.

4. PUTTING IT ALL TOGETHER – YEARLY STREAMFLOW ANALYSIS IN SICILY

The previous sections have introduced the methods that will be used in the analysis of yearly streamflow in Sicily. It is Mediterranean's largest island and falls between the 36° and the 38° parallel north and has an area of 25400 km^2. It is mostly hilly and mountainous. In addition to Etna's imposing volcanic cone (3370 m above sea level), Sicily has four main mountain groups. The first (the Sicilian Apennines) stretches along the north coast, from the Strait of Messina to the Torto River, as a continuation of the Calabrian Apennines and its highest peaks rise to about 2000 m. The second group (Sicani) encircles the western part of Sicily to the west of the Torto and Platani rivers. The third mountain group forms the heart of the island, and overlooks the African Sea to the south-west; its most characteristic part is often referred to as the "sulphur-bearing upland". The south-east corner of Sicily is mostly characterized by a plateau (Hyblaean Mountains). The island has little flat land: the largest stretch, 430 km^2, is the Catania Plain, between Etna and the mountains of the province of Syracuse. Other flat areas lie in the province of Trapani, near Marsala, Mazara and Castelvetrano. The major rivers are the Simeto (88 km, with a river basin of 4186 km^2), the Salso (or South Imera – 112 km with a river basin of 2200 km^2) and the Platani (84 km with a river basin of 1784 km^2), but they are almost dry in summer. The climate is typically Mediterranean along the coasts, with hot but not torrid summers, mild and short winters, and moderate rainfall (from October to March). The annual average of clear days is 98 in Palermo; 110 in Messina; 130 in Taormina; 133 in Syracuse. The annual average temperature along the coast is between 17° C and 18.7° C, July being the hottest month. The typical evergreen Mediterranean scrub vegetation is widespread. There are still traces of the oak-woods which must have covered the lower mountains in ancient times, as well as of the beech-woods which form the upper belt of the Nebrodi and Madonie woods.

4.1. The data base

Daily records of streamflow are available at more than 50 sites from the UIR (Ufficio Idrografico Regionale – Sicilian Hydrographic Office). In this study, 43 records with a length of at least ten years in the period 1971 – 1997 were employed, even with missing years. Year 1997 was the last when streamflow data have been published. Yearly streamflow totals (in mm) have been used in the frequency analysis and their at-site sample mean has been used as index. Records longer than ten years but prior to 1971 have not been included in the analysis in order to account for land use change, more intensive exploitation of groundwater resources and other factors which, together with the likely reduction of precipitation amounts in this area (Cannarozzo *et al*, 2006), make streamflow records dating back to, say, the '30s of the last century hardly reliable for the assessment of water resources. In addition, 20 streamflow records at various dams on the island have been used that have been obtained by means of the volume budget of the reservoir at monthly scale. Inflow records at dams constitute a precious source of knowledge because, unlike streamflow records, they span over the last ten years. As they have been obtained using different methods, the Kruskal-Wallis test (e.g. Hirsch et al., 1993) was used to assess the homogeneity of two such groups of records. Table 1 reports the results of the Kruscal-Wallis test for average annual streamflow and for mean annual runoff ratio. It shows that the value of the KW statistic falls within the region of acceptance of the null hypothesis, so that gauged and reconstructed runoff can be treated as a single sample. The average mean yearly runoff per unit area, together with its maximum and minimum value and its estimated variability over the island is reported in table 2. The reported average mean yearly runoff is the weighted (by the basin's area) average of mean yearly streamflow.

Table 1. Results of the Kruskal Wallis for homogeneity testing of gauged and reconstructed (via reservoir water balance) annual flows

Sample size		KW		
UIR stations	Reservoirs	Mean annual flow	Mean annual runoff ratio	$\chi^2_{c,\,0.05}$
43	20	1.69	2.09	3.84

4.1.1. Explanatory variables

A number of morphologic and climatic variables have been proposed in many studies to explain the variability of mean annual streamflow over an area (e.g. Orsborn, 1974). Many of these variables, however, do not prove significant in explaining the variability of average annual runoff. Some form of rationale is necessary to identify geomorphic and climatic variables that can be related to the long term water balance at catchment's scale. Plaut Berger & Enthekabi (2001) have identified six climatic and physiographic variables that are conceptually related to basin-scale equilibrium hydrology. The selection of the following four variables is in part motivated by their results and by the ensuing comment by Sankarasubramanian & Vogel (2002):

- Mean yearly rainfall P[mm];
- Mean basin elevation H [m a.s.l.];

- Mean basin slope S [%];
- Basin drainage density D [km/km^2];

Rainfall: for a given basin, mean yearly rainfall is computed by the Thyessen method using yearly rainfall totals recorded at the UIR network. The figures of total rainfall supplied by the Thyessen method can be rather coarse, especially when orography plays a significant role on the precipitation field and few or no rain gauges at high elevations are available, as it is the case in the South-Eastern part of the island. In the northern part of the island the relationship mean annual precipitation – mean basin elevation is weaker, as the result of the proximity of relief to the coast (the Apennine chain is sub-parallel and close to the coast). Overall, as shown in table 2, the average mean yearly rainfall over the selected basins is 630 mm with a maximum of 1043 mm and a minimum of 443 mm.
Mean basin elevation: mean basin elevation of the selected basins was obtained via a Digital Elevation Model (DEM) of the island with a 40x40 m grid resolution.
Mean basin slope: a slope map of the island was derived from its DEM using the "Derive Slope" function of the Arcgis software. The function "raster calculator" has then been used to obtain the distribution of slopes and its average value over a given basin.
Drainage Density: Drainage density is defined as the total length of the channel in the basin per unit area and depends on the resolution of the map and on the order of the channels. The map of the channels was supplied by the UIR and was obtained by ortophoto map digitalization at scale 1:10.000. The map shows two different levels of accuracy in drawing the channels. As this has an impact on the total channel length and ultimately on the significance of such variable, we have scaled the drainage density by the total number of channels in the basin. Such variable will be denoted in the following with D'd.
Regression of mean annual streamflow on such variables yields a R^2 of 0.77 but only precipitation, mean elevation and slope prove significant according to a t test with 5% significance with a R^2 of 0.74 (table 3). According to the value of standardized Moran's I for the residuals of such regression, however, (table 4) there is a significant spatial autocorrelation among residuals, suggesting the existence of missing covariates. We then added the following variables:

- Thornthwaite wetness index I_T;
- A base flow index BF [litres/(s*km^2)];
- North/south dummy.

Thornthwaite wetness index: The Thornthwaite wetness Index (Thornthwaite, 1948) is a popular index summarizing the soil water balance at yearly scale at a given site. It is defined as $I_T = (P – Et_p)/Et_p$, where P is the mean yearly rainfall total and Et_p is the yearly mean evapotranspiration. I_T can also be used as a climatic index and as such it has been mapped for Sicily (Regione Siciliana, 2004). In Regione Siciliana (2004), I_T was arranged in six classes to provide a climatic index. A map is available depicting such climate classification according to I_T, ranging from hyperhumid ($I_T > 1$) to arid ($I_T < -0.67$). Classes are labelled from 1 (arid) to 6 (hyperumid). We worked on such labels (from 1 to 6) rather than on the actual values of the Thornthwaite index using ARCGIS tools to obtain an average (non integer) climatic index of the basin.
Base flow index: Many Sicilian basins are ephemeral: the water balance closes within the water year (approximately October – May) as there is very little or no baseflow at all. Most basins have a quick response to precipitation and streamflow decays rapidly after peak floods. Some basins, however, have perennial or semiperennial features with longer recession periods and streamflow also in summer. In a long-term, average perspective as the one of this paper, interactions between surface and groundwater should play a limited role, as in general the main effect of infiltration and groundwater flow is to lag and modulate the response to the input (the rainfall). This statement is likely to hold if the basin is big enough to contain entirely one or more aquifers. However, as Sicilian basins are usually quite small, a further variable allowing for groundwater flow from outside the watershed may be useful to describe the interactions between aquifer and watershed in basins with some baseflow as different watersheds lay on the same aquifer. Groundwater supply to the basin from outside its boundaries usually comes through springs. Intuitively, the larger the number of springs in a watershed, the likelier the occurrence of this kind of mechanism. A variable is then introduced BF = Total estimated spring yield per unit area (litres/s*Km^2). As naïve as this variable may appear, since some kind of double-counting of water resources could occur if springs in a basin are supplied by an aquifer entirely included within the watershed, it seems to be able to capture the phenomenon, as suggested by the improvement in the percentage of explained variance from 0.80 to 0.85 (see table 2) and its significance to the t test.The spring yield was assessed using the to-date most complete inventory of springs in Sicily (with measures) dating back to the 1930s. Flow values may have radically changed and some springs may even no longer exist: what remains unaltered (or almost unaltered) in

comparative terms (i.e. among different basins) is the different degree of ability of the basin to supply the river through springs.

North/South dummy: this variable was introduced to account for a "hillside effect" due to the mountain range along the island's northern coast which produces different average amounts of precipitation at the same altitude on its southern and northern sides. Basins with their mouth in the Tyrrhenian Sea (north) have been labeled with 1, basins with their mouth in the Ionian Sea (east) or on the Sicilian Channel (south) have been labelled with 0. Overall, this variable proves significant to the t-test with a 5% significance level.

4.2. Regression analysis for mean annual streamflow

In many environmental applications, the covariates explaining a certain process are not mutually independent; rather, they frequently exhibit correlation, in contrast with the assumption of independence among explanatory variables of the classic regression model. The effects of near multicollinearity are rated differently in the literature (Chung-Ming Kuan, 2002): although in the absence of exact multicollinearity parameters can still be estimated by the OLS method and they preserve all the properties of OLS estimators, yet the variance of the parameter estimates may be very high so that t-ratios are likely to be insignificant. VIF (Variance Inflation factor) is a measure of such risk. It is defined as $VIF_i = (1 - R_{VIF}^2)^{-1}$ where R_{VIF}^2 is the R^2 of the linear regression of the i-th explanatory variable on all the others. A rule of thumb is that correlation among the i-th variable and the others should not be an issue if $VIF_i < 5$. This is the case in this application, where the highest VIF is 4.5 (table 3).

Table 4 contains the values of standardized Moran's I for the residuals of the regression of mean annual discharge against the above explanatory variables using the entire sample of 63 basins. Results are reported with reference to groups of different explanatory variables to show the effect on results of the stepwise introduction of I_T, BF and the north/south dummy in the regression. While the residuals of regression on P, Hm and S are autocorrelated in space, suggesting the existence of further covariates, the low value of the index for the other regressions indicates that the residuals can be regarded as an independent spatial process.

Table 2. Basic statistics of the explanatory variables of mean annual streamflow

	Mean annual streamflow	Mean annual rainfall	Average Basin Elevation	Average Basin Slope	Average scaled Drainage Density	Thornthwaite's climate index	Base flow index	Area
	[mm]	[mm]	[m above sea level]	[%]	[km/km^2]		[litres/(s*km^2)]	[km^2]
Mean	137.0	630.0	668.5	20.9	1.1	3.2	1.0	173.8
Min	31.0	442.7	113.0	1.2	0.4	1.9	0.0	8.8
Max	845.3	1043.0	1479.0	40.6	3.0	5.5	9.0	1789.0
Standard deviation	174.1	132.1	309.0	8.3	0.5	0.9	1.8	279.8

On the other hand, results of the White test on the same residuals, reported in table 3, show that heterogeneity exists and the assumption of constant variance for the whole sample does not hold: the test statistic W (first row on the left-hand side of table 3) is greater than the critical $\chi^2_{0.01}$ value for each of the different combinations of the explanatory variables, suggesting that some correlation does exist between the residuals variance and the explanatory variables. Table 3 also shows the results of the Goldfeld Quandt test confirming that heteroscedasticity exists. Interestingly enough, the highest G.Q. values are obtained in all cases for climatic variables such as rainfall or Thornthwaite's climatic index while no heterogeneity seems to stem from the climate-independent baseflow covariate (G.Q. on B.F. = 1.0). This suggests splitting the sample in two subsamples according to some climatic criterion. Probably, the simplest way to do this is to select an average or median value of P or I_T and label as "subhumid" and "semiarid" those basins with P or I_T over and below such average value. We opted for a classification based on I_T as it contains more information than the plain rainfall datum and for the median rather than for the mean as it is less influenced by outliers. The median value of the I_T sample is 3.02. As showed in table 3, the effect of splitting the sample in two subsamples according the above climatic criterion is to dramatically reduce heteroscedasticity, so that according to the White test no further subdivision is required for both semiarid and subhumid basins, whereas the Goldfeld-Quandt test suggests that mean basin elevation (and average basin slope in subhumid basins) could still be a potential source of non-constant variance in the regression's residuals.

Table 3. Results of the White and Goldfeld-Quandt test and regression statistics for the regional model of mean annual streamflow

	WHITE TEST				GOLFELD QUANDT TEST					
Significant variables	P,Hm,S	P,Hm,S,I_T	P,Hm,S,I_T,BF	P,Hm,S,I_T,BF, N/S	P,Hm,S (F value(0.05)= 1.84)		P,Hm,S,I_T (Fvalue(0.05) = 1.86)		P,Hm,S,I_T,BF,N/S (Fvalue(0.05) = 1.88)	
W = nR^2	34.9	31.4	39.0	41.4	G.Q.on P	p-level	G.Q. on P	p-level	G.Q. on P	p-level
Degrees of freedom	9	14	20	21	13.4	2.7E-07	11.8	1.5E-06	5.3	1.2E-04
$\chi^2_{d.o.f.,0.01}$	16.9	23.7	37.6	38.9	G.Q. on Hm	p-level	G.Q. on Hm	p-level	G.Q. on Hm	p-level
p level	6E-05	0.005	0.007	0.005	4.6	8.4E-04	8.7	1.5E-05	3.4	4.0E-03
	REGRESSION ANALYSIS				G.Q. on S	p-level	G.Q. on S	p-level	G.Q. on S	p-level
R^2	0.74	0.81	0.86	0.87	4.5	9.5E-04	5.3	4.3E-04	4.1	1.4E-03
Adjusted R^2	0.73	0.80	0.85	0.85			G.Q. on I_T	p-level	G.Q. on I_T	p-level
RMSE [mm]	90.9	79.4	68.3	66.9			11.8	1.5E-06	8.5	2.2E-06
Mean/RMSE	2.2	2.6	3.0	3.0					G.Q. on BF	p-level
Max VIF	2.2	3.5	3.6	4.5					1.0	5.1E-01

In order to account for such indications from the tests and considering that the "hillside" effect proves significant both at regional scale and for semiarid basins, we further split the sample of semiarid basins in two subsamples (north/south) and resort to some operational criterion for subdividing semihumid basins that may be helpful to increase the likelihood of the homoscedasticity assumption while also hopefully improving prediction of mean yearly streamflow in subhumid basins. The criterion could be to split the sample of semihumid basins so to minimize the root mean squared error (RMSE) of the two regressions subject to a proximity constraint, i.e. the two groups must be formed by neighbouring basins. To do this, the sample of 32 subhumid basins was first ordered according to a geographical criterion (from north east counter clockwise).

Table 4. Moran's I and standardized Moran's I for different sets of covariates. For N = 63 and for the given spatial dependence structure W, E[I] = -0.016 and VAR[I] = 0.0035

Significant explanatory variables	P,Hm,S	P,Hm,S,I_T	P,Hm,S,I_T,BF	P,Hm,S,I_T,BF, N/S
Moran's I	0.129	0.009	0.047	0.049
Standardized I	2.453	0.437	1.089	1.122
p-level	0.007	0.331	0.138	1.124

Starting from basin ranked 1 according to the above geographical criterion, two groups were formed, one with n1 and the other with 32 – n1 elements, with n1 varying from 7 to 32 – 7, thus obtaining 25 pairs of groups and for each pair the total RMSE was computed. The number 7 is the maximum number of explanatory variables plus 1. Then, the same procedure is repeated starting from the basin ranked 2 (rank 1 basin is now at the end of the list) and other 25 pairs of groups are obtained and

RMSE is computed for each of such pairs, then the basin ranked 3 is taken as starting point (basin 2 is now at the end of the list preceded by basin 1) etc.
It turns out that the minimum RMSE is obtained for two groups roughly subdividing the island along a north-south direction, so that the two subgroups may be denoted as "Subhumid west" and "subhumid east". Results of the tests on the regressions residuals (data not reported) confirm that for the four subsamples the assumption of homoskedasticity is now supported by both tests. The final relationships linking mean annual streamflow to the climatic and physiographic features of the basins are reported in table 6.
With the exception of basins belonging to the "Subhumid west" typology, the regression models seem to provide acceptable results, as confirmed by the $R^2_{\text{cross validation}}$ values. Cross validation is a simple way to assess the ability of the regression model to predict mean annual streamflow in ungauged basins (that cannot hence have been used for calibration) or, seen differently, a way to assess the influence of the data on results.

Table 5. Results of the White and Goldfeld-Quandt test and regression statistics for the two models of mean annual streamflow in semiarid and subhumid regimes

	WHITE TEST		GOLFELD QUANDT TEST			
Type of catchment	SEMIARID	SUBHUMID	SEMIARID		SUBHUMID	
Significant variables	P,Hm,BF,N/S	P,Hm,S,I_T,BF	P,Hm,BF,N/S (F value(0.05) = 4.3)		P,Hm,S,I_T;BF (Fvalue(0.05) = 4.3)	
nR^2	12.6	25.2	G.Q.on P	p-level	G.Q. on P	p-level
Degrees of freedom	12	20.0	2.31	0.31	2.39	0.11
$\chi^2_{\text{d.o.f.},0.01}$	26.2	37.6	G.Q. on Hm	p-level	G.Q. on Hm	p-level
p level	0.40	0.19	4.78	0.01	6.50	0.01
	REGRESSION ANALYSIS		G.Q. on BF	p-level	G.Q. on S	p-level
R^2	0.85	0.86	1.70	0.22	4.75	0.15
Adjusted R^2	0.84	0.83			G.Q. on I_T	p-level
RMSE [mm]	26.56	80.0			1.52	0.27
Mean/RMSE	3.8	3.9			G.Q. on BF	
$R^2_{\text{cross validation}}$	0.81	0.75			0.71	0.69

Here it is performed by leaving out one station each time, calibrating the model with the remaining stations and calculating the MAF for the missing station (and hence the error) via the calibrated model. The $R^2_{\text{cross validation}}$ is given by $1 - \sigma^2_\varepsilon/\sigma^2_{MAF}$, being σ^2_ε the variance of errors and σ^2_{MAF} the MAF sample variance.

Table 6 Regression models of mean yearly streamflow and regression statistics for the four basin typologies

Type of basin	Sample size	Equation	R^2	R^2_{adj}	$R^2_{\text{cross validation}}$	RMSE [mm]	Mean/ RMSE	max VIF
Semiarid north	10	MAF = -133.83 + 0.43*P	0.74	0.74	0.66	14.6	6.5	-
Subhumid west	21	MAF = -319.33 + 0.51*P+ 0.15*Hm + 54.24*I_T - 7.13*S + 18.05*BF	0.73	0.66	0.38	44.1	5.0	4.0
Semiarid south	22	MAF = -143.82 + 0.22*P + 0.16*Hm + 41.78*BF	0.89	0.87	0.85	27.5	3.8	1.2
Subhumid east	10	MAF = -447.32 + 0.48*P + 98.48*I_T	0.96	0.96	0.93	49.3	10.1	2.8
REGION	63		0.97	0.96	0.94	42.9	4.7	-

Three out of four of such models seem to be robust, in that the available data influence less the results, as showed by the similarity of the adjusted R^2 and of the $R^2_{\text{cross validation}}$. However, the poor performance of the subhumid west basins is to a good degree to be ascribed to a single station without which the $R^2_{cross\ validation}$ rises to 0.53. It is maybe worthwhile observing that the introduction of a further subdivision in subhumid basins improves the overall performances of the models, with an adjusted R^2 (the adjustment is necessary to allow for the increased number of parameters, as the four models now have a total of 15 parameters) of 0.96 and a $R^2_{cross\ validation}$ of 0.94. For the semiarid/subhumid scheme at a regional level R^2_{adj} = 0.90 and $R^2_{cross\ validation}$ = 0.85.

4.3. Growth curve homogeneity analysis

The procedure outlined in section 3 was applied to the available sample of indexed annual streamflows to identify regions that can be assumed to have the same growth curve. 1116 station-years have been employed in this study. The Hosking-Wallis procedure was first applied to the whole sample and then to the same subsamples (semiarid and subhumid basins) that were used for the analysis of mean annual streamflow. Figure 1 shows the pattern of the sample L-cvs for semiarid and semihumid basins. As was expected, the average L-cv of the semiarid basins is greater than its counterpart in semihumid ones, as a consequence of the greater variability of yearly streamflows in arid and semiarid basins. Results of the Hosking-Wallis procedure are reported in table 7. It shows that Sicily cannot be considered an homogeneous area from the standpoint of the growth curve of annual streamflow. On the other hand, basins meeting the criterion I_T<3.02 and hence classified as semiarid have a negative H statistic, implying that the simulated variability of the L-cvs exceeds that of the sample. This may be due to the poor fit to data of the tails of the kappa distribution: as H is built on V, the sample (simulated and real world) L-cv variance, itself a notoriously outlier-sensitive product moment, any departure in the tails between data and fitted distribution may produce larger variability of the simulated L-cvs, hence giving rise to negative H values.

Table 7. Results of the Hosking-Wallis procedure for the identification of homogeneous regions

	Sicily	Semiarid	Semihumid	Subhumid			
				East		West	
				Overall	without outlier	Overall	without outlier
Average L-Cv	0.341	0.420	0.248	0.240	0.200	0.286	0.273
V_{oss}	0.02	0.007	0.004	0.009	0.002	0.004	0.003
H	6.54	-1.227	3.828	9.920	0.555	1.750	0.300

Clearly, it could be questioned if the assumption that the sample has a kappa distribution as parent distribution should be kept in such circumstances and if the ensuing conclusions, i.e. that the sample is homogeneous, may be accepted. The assumption of homogeneity is provisionally accepted, although further analysis may be developed, for instance using a non parametric test, such as the bootstrap Anderson Darling test introduced by Viglione et al.(2007) that seems to be superior to the HW procedure in regions with high skewness, such as semiarid areas. Subhumid basins instead are heterogeneous, as pointed out by their H value. Heterogeneity does not seem to be reduced, at least at first glance, by splitting the sample in two subsamples in the same fashion as in the previous section ($H_{semihumid-west}$ = 1.750, $H_{semihumid-east}$ = 9.920). Such high H values however may be due to the presence of outliers. As a matter of fact, leaving out the largest L-cv value (L-cv= 0.50) from the sample of the subhumid western basins and the largest L-cv value (L-cv= 0.554) from the eastern basins, considerably lowers the value of V_{obs} and the corresponding H value. It may be worthwhile observing that the two L-cv outliers correspond to I_T values (3.021 and 3.024 respectively) that are located on the frontier between the two climate regimes.

4.4. Parameter estimation and goodness of fit of some candidate distributions

The frequency analysis of annual streamflow is completed by the selection of an appropriate growth

curve, one for each of the three above homogeneous areas. The three indexed samples are skewed with asymmetry coefficients γ of 2.432 for semiarid basins and 1.03 and 0.845 for eastern and western subhumid basins respectively. Such figures are consistent, in that flow characteristics are usually more skewed in semiarid basins than in subhumid ones. A number of pdfs have been screened and three types, namely a Generalized Extreme Value (GEV), a three parameter lognormal (LN3) and a Pearson type 3 (P3) distribution have been selected for further analysis. Parameters of the GEV distribution have been estimated via the method of L-moments (Stedinger et al,1993).
The lower bound parameter of the LN3 has been estimated through a quantile-lower-bound estimator (e.g. Stedinger et al., 1993). Finally, the P3 parameters have been estimated through a simple least square (Martin, 1999) procedure for the evaluation of the lower bound term and the other two parameters have been estimated by the method-of-moment relationships between them and the lower bound parameter.
Goodness of fit has been assessed using the Anderson-Darling test, an EDF (Empirical Distribution Function) test that quantifies the departure of the theoretical cumulative distribution function (CDF) $F(x,\boldsymbol{\theta})$, being $\boldsymbol{\theta}$ the parameter vector, from the empirical CDF that is obtained from the sample of size n by ranking its elements in ascending order and letting $F_n(x_i) = 1$ if $x_i > x_n$ and $F_n(x_i) = i/n$ elsewhere. The Anderson Darling statistic is defined as follows:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)\ln[F(x_{(i)},\theta)] + (2n+1-2i)\ln[1-F(x_{(i)},\theta)]\right\} \tag{7}$$

To perform the test, one needs to know the distribution of A^2 under the null hypothesis H_0, i.e., when the sample is distributed according to $F(x,\boldsymbol{\theta})$. When there is no a priori knowledge of the parameters but they must be inferred from the data, as it is the case in this and in most applications, then the quantiles of the statistic depend in general terms on the shape of F, on the sample size and even on the estimation procedure employed (Stephens' case p, Stephens (1986)).
In this paper we resort to simulation (Viglione *et al.*, 2006)) to obtain $G(A^2_0)$, the cumulative distribution function of A^2 under the null hypothesis. Given $F(x,\boldsymbol{\theta})$, a large number (1000 in this application) of samples of the same size as the historic one are extracted from the distribution. In order to allow for the p-case, parameters $\hat{\theta}$ of a distribution of the same type as F are estimated for each generated sample and a distribution with such parameters is used to assess $F(x_{(i)}, \boldsymbol{\theta})$ in (7). The A_0^2 values are ranked to form an empirical cumulative distribution function which is assumed to be the actual distribution of A^2 under H_0. Using a quantile $1-\alpha$ (e.g. $\alpha = 0.10$; 0.05 or 0.01) as upper bound of the acceptance region, if the sample A^2 has a value such that $G(A^2) > 1-\alpha$, then the hypothesis that the sample is distributed according to the $F(x,\theta)$ must be rejected at the selected significance level.
Table 8 contains the results of the test. The P3 certainly supplies the poorest results, with the data of no region being fitted; however, no of the two other candidate distributions can be said to fit the data in all the regions. In the subhumid West region, only the LN3 proved acceptable at the 1% level. Not surprisingly, this is the area with the (by far) lowest cross-validation R^2 in the evaluation of the mean annual flow (table 6), which once again suggests the close relationship between regional analysis of mean annual streamflow and of its pdf. However, as the GEV passes the test at a 5% significance level in two out of the three regions, it is selected for the semiarid and subhumid basins of the eastern part of the island, while a LN3 is preferred for the western subhumid basins.
In the following, the expressions of the growth curve for the three areas using a GEV and the adopted LN3 for the subhumid western basins are provided.

Semiarid: $x_p = 0.598 - 1.634*\{1 - [-\ln(p)]^{0.265}\}$ (8a)

Subhumid west $x_p = 0.796 + 4.578*\{1 - [-\ln(p)]^{0.090}\}$ (8b)

Subhumid east $x_p = 0.844 - 3.714*\{1 - [-\ln(p)]^{0.084}\}$ (8c)

Subhumid west LN3 :$x_p = \ln \Phi^{-1}(p, 0.713, 0.235, -1.099)$ (8d)

Where Φ^{-1}denotes the inverse of the normal distribution and the three numbers witin brackets are the estimated mean, standard deviation and lower bound.

5. CONCLUSIONS

Annual streamflow frequency analysis has been performed in Sicily using the index flow procedure. We found that heterogeneity both at the level of the index (mean annual streamflow in this application) and of the growth curve can be ascribed to different climate regimes existing in the island, due to its orography. In a spatial framework, the violation of each of the assumptions of the classic linear regression model can provide information on the process itself: while autocorrelation among residuals may be the signal of an omitted covariate or of the existence of small-scale processes, heteroscedasticity is the symptom of the existence of different levels of the output that may need different parameters (and probably different covariates) to be explained. Using a popular climate index such as that of Thornthwaite's and some considerations based on the results of tests on heteroscedasticity, we could identify two such regimes that were further subdivided based on operational considerations.

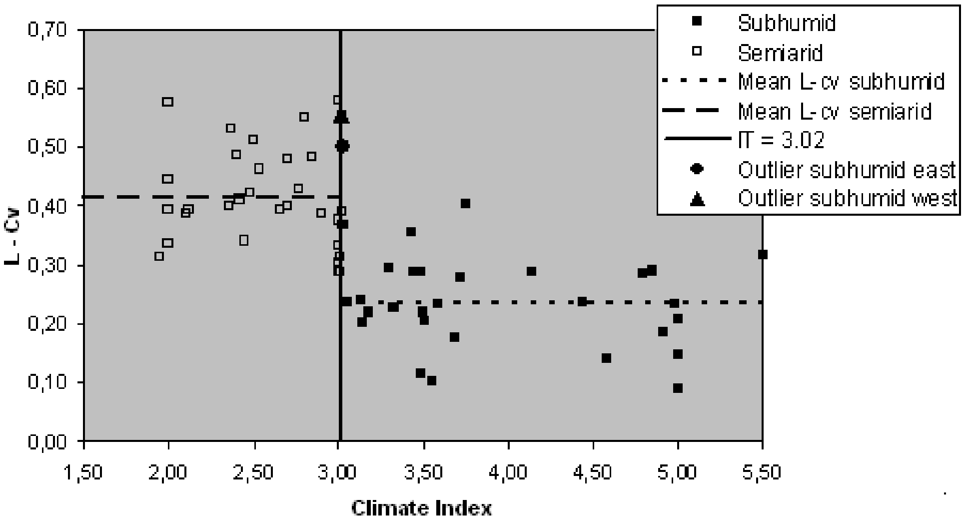

Figure 1 Sample L-cvs for subhumid and semiarid basins, their mean values and the frontier (I_T = 3.02) between the two climate regimes.

Table 8. Results of the Anderson-Darling test for GEV, LN3 and P3 distributions

	GEV		LN3		P3	
	A^2	$P(A^2)$	A^2	$P(A^2)$	A^2	$P(A^2)$
Semiarid	0.563	0.91	1.348	0.888	3.682	1
Subhumid West	1.307	0.998	1.074	0.971	1.535	1
Subhumid East	1.362	0.659	1.555	0.998	1.690	1

Overall, the efficiency of the four regression models is quite good; however, poorer predictions are expected in one of the subareas. Significantly enough, such identified regions also prove, with some minor adjustments through the elimination of two outliers in the two subhumid regions, homogeneous from the standpoint of the pdf of the indexed annual flows. The analysis is completed by the selection of a growth curve. A GEV is selected for two of the subareas and a three parameter lognormal for the other. Besides further inspection of the subregion where both regression analysis of mean annual streamflow and goodness of fit evaluations yielded the poorest results, some methodological questions remain open: in the first place, although cross-validation provides some kind of assessment of the capability of the models to predict mean annual streamflow at ungauged sites, more data for calibration should be needed both for mean annual flow and for quantile assessment.To this end, some ideas from spatial experimental design may be used to reduce the number of stations to use for calibration and to use the remaining records for further validation; in the second place the role of uncertainty in both parameter estimation and in the data is somewhat left in the background and needs to be quantified. While uncertainty about the model and the data can be somehow incorporated in the error structure via alternative estimation procedures (such as generalized least squares), uncertainty on the future average rainfall levels or the changing distribution of the spatial regimes over the island due to climate change may be accounted for by different scenarios based on the outcomes of climate change studies.

6. REFERENCES

Cannarozzo, M., Noto, L.V., Viola, F. (2006). *Spatial distribution of raifall trends in Sicily* (1921-2000) Physics and Chemistry of the Earth, 31, 1201-1211.

Cliff, A.D. and Ord, J.K. (1981). *Spatial Autocorrelation*, Pion, London.

Cressie, N. (1993). *Statistics for spatial data*, John Wiley and Sons, New York

Darlymple, T. (1960). *Flood frequency analyses*, US Geol. Surv. Water Supply paper 1543-A.

Hirsch, R.M., Helsel, D.R., Cohn,T.A. and Gilroy, E.J. (1993). *Statistical analysis of hydrologic data* in Maidment, R.D., "Handbook of Hydrology", Mc Graw – Hill, New York, USA.

Hosking, J.R.M. (1994). *The four parameter kappa distribution*, IBM J.Res.Develop. 38(3).

Hosking, J.R.M. and Wallis J.R. (1993). *Some statistics useful in regional frequency analysis*, Water Resources Research, 29(2), 271-281.

Hosking, J.R.M. and Wallis J.R. (1997). *Regional Frequency analysis: An Approach Based on L-moments*, 1st ed., Cambridge University Press, New York

Kuan, C.M. (2002) *Econometrics* http://www.sinica.edu.tw/as/ssrc/ckuan/index-e.html

Martin D. (1999), *Estimation of mean rain rate through censoring*, J. Appl. Meteor., 38(6), 797-805.

Orsborn, J.F. (1974). *Determining streamflows from geomorphic parameters*, J. Irrig. Drain. Div. ASCE, 11 (IR4), 455-475.

Plaut-Berger, K., and Entekhabi, D. (2001), *Basin hydrologic response relations to distributed physiographic descriptors and climate, Journal of Hydrology*, 247, 169-182.

Regione Siciliana - Assessorato Agricoltura e Foreste – Servizio Informativo Agrometeorologico Siciliano (2002), *Atlante Climatologico della Sicilia II Edizione*, Palermo,Italy (available in CD ROM)

Sankarasubramanian, A. and Vogel, R.M. (2002). *Comment on: "Basin hydrologic response relations to distributed physiographic descriptors and climate" by Karen Plaut Berger, Dara Enthekaby, 2001.Journal of Hydrology, 247, 169-182*, Journal of Hydrology, 263(1), 257-261.

Stedinger, J.R., Vogel, R.M. and Foufoula-Geourgiou, E. (1993), *Frequency analysis of extreme events* in Maidment, R.D., "Handbook of Hydrology", Mc Graw – Hill, New York, USA.

Stephens M.A.(1986), *Tests based on EDF statistics* in D'Agostino R.B., and Stephens M.A. "Goodness of Fit techniques", eds, Marcel Dekker, New York.

Thornthwaite, C.W., (1948), *An approach toward a rational classification of climate*, Geophysical Revue 38, 55 – 94.

Viglione A., Claps P., and Laio F. (2006), *Utilizzo di criteri di prossimità nell'analisi regionale del deflusso annuo*, XXX Convegno di Idraulica e Costruzioni Idrauliche – IDRA 2006, Rome.

Vogel R.M., Wilson I. and Daly C., (1999), *Regional regression models of annual streamflow for the United States*, J. Irr. Dr. ASCE, 125(3), 148 – 157.