European Water 35: 31-44, 2011.

Geostatistical Techniques for Runoff Mapping: An Application to Sicily, Italy

A. Di Piazza, D. Caracciolo, L.V. Noto, F. Viola and G. La Loggia
Dipartimento di Ingegneria Civile, Ambientale e Aerospaziale, Università degli Studi di Palermo, Italia,
dipiazza@idra.unipa.it, caracciolo@idra.unipa.it, valerio@idra.unipa.it, viola@idra.unipa.it, glal@idra.unipa.it

Abstract: The availability of reliable and long time series of runoff data is fundamental for most of the hydrological analyses and for the assessment and the management of water resources even in condition of global climatic change. However, hydrologic data sets are often characterized by a short duration and also suffer from missing data values, mainly due to malfunctioning of gauging stations for a specific period. In order to overcome this problem and obtain long and continuous runoff time series, different models and methods have been previously developed and proposed. While some models, used to extent the streamflow record, are conceptual, empirical, regressive models based on the rainfall input, other models are based on the derivation of runoff maps at different time scale; these maps allow the runoff estimation in gauged basins characterized by the absence of data or in ungauged basins. Aim of this paper is the derivation of a map relative to the mean annual runoff at regional scale using a stochastic approach derived from the kriging interpolator. This approach can be assimilated to a kriging system, which considers explicitly the areal nature of runoff variable by imposing the constraint of the water balance; it allows to derive gridded annual runoff maps with finer and finer resolution. The methodology has been applied to 23 main hydrographic basins of Sicily, Italy using the mean annual runoff dataset provided by Osservatorio delle Acque. All these basins have been previously grouped in three homogeneous zones, as suggested by previous studies. A cross-validation procedure has been performed in order to validate the procedure for each homogeneous zone.

Key words: runoff, nested basins structure, geostatistical methods, hierarchical approach, ungauged basins, gridded maps.

1. INTRODUCTION

The reconstruction of incomplete hydrologic data records has been the subject of a large number of scientific works where numerous techniques for estimating missing data values have been implemented and compared.

Many hydrological researchers have adopted and developed various models and techniques to deal with the problem of estimating missing data. The efforts are devoted not only to extending short records by adding lengthy segments of estimated data, but also attention is given to the short duration gaps filling.

The methods used in literature to estimate the missing values in runoff time series, can be classified as: (1) rainfall-runoff models (Maskey, 2009) and (2) methods for the derivation of runoff maps. The first class can be subdivided into (1a) physically-based, (1b) conceptual and empirical and (1c) data-driven or black box (neural network, fuzzy algorithms, etc.) (Ilunga and Stephenson, 2005; Cutore et al., 2006) while the second class contains (2a) subjective methods (i.e. manual contouring maps; Arnell, 1995) and (2b) stochastic methods. (Gottschalk et al., 2006; Skøien and Bloschl, 2006).

Moreover, all these models and methods can be distinguished into the following three general classes models, as a function of their application domain: (i) spatial models, which represent the spatial distribution of variables over a specific duration, (ii) temporal models, which represent the variables at a point over time and (iii) space-time models, which represent both the spatial and temporal evolution of the variables.

Among these methods, particular attention has been paid during the two last decades to the interesting methods for the derivation of runoff maps (i.e. deterministic or stochastic interpolation techniques). Through these maps it is possible estimate runoff in a certain region even if interpolation of runoff is more complex than interpolation of the variables usually assimilated to a point process since it is a generalized random space-time process with a local support equal to the basin area.

There are three main issues to be considered when choosing methods for the construction of runoff maps (Gottschalk and Krasovskaia, 1998): a) the method to be used for interpolation, b) the scale of fundamental units on the map, and c) the available observations that can be used to investigate on runoff variability at different spatial scales. The interpolation method, as mentioned above, can be either manual contouring, or automatic interpolation (deterministic or stochastic) based usually on a weighted average. The second topic that needs attention is the scale. On meso and micro scales, the area of drainage basins needs to be taken into account in the interpolation procedure, which has several advantages compared to the point interpolation. When basins are considered as points in a continuous space, the lateral aspects of the runoff process are neglected. Therefore, one cannot expect that runoff in this case, when integrated over a river basin, coincides with measured streamflow in the main rivers. The third topic to be considered is the type of available observations to resolve the variability across space at different scales. The estimated spatial variability from a regional set of observations can be expected to depend on the size of the basins involved (i.e. the higher the variability, the smaller the basins).

A hierarchical approach for interpolation is elaborated by Sauquet et al. (2000). The territory to be mapped is divided into sub-basins in a hierarchy of scales. The number of levels in this hierarchy is determined mainly by the amount of available observations, which also indicates the level of detail that can be achieved. The first level in a larger drainage basin is usually already well defined by existing gauging stations in the main rivers constituting the first level of sub-basins. These basins are, in turn, divided into a second level of sub-basins (or grid cells), and gauging stations with appropriate basin scales are chosen as the background for the interpolation. The interpolation procedure guarantees that the water balance equation is satisfied so that the sum of runoff from this second level of basins is equal to that of the first order basin including them. The procedure can be repeated to a third level and so on. At each step new information must be added.

Skøien and Bloschl (2005) proposed the Top-kriging, or topological kriging, as a method for estimating streamflow-related variables in ungauged catchments. The main appeal of the method is that it is a best linear unbiased estimator (BLUE) adapted for the case of stream networks without any additional assumptions. This method of geostatistical estimation on stream networks extends the original work of Sauquet et al. (2000) in a number of ways. First, they suggested that the interpolation method can be used, in an approximate way, for a range of runoff related variables including variables that are not fully mass conserving. Second, they used variograms while Sauquet et al. (2000) used covariances, allowing to deal with variables that are non-stationary. Third, they accounted for local uncertainties of the measurements that may differ between locations. Last, they illustrated the potential of the approach for estimating the uncertainty of the variable of interest in ungauged catchments. This approach was applied for estimating the 100 year specific flood in ungauged catchments in Austria. Other studies about the development of this method have been carried out, in particular deepening the issues relative to correlation and covariance of runoff and distance measures for hydrological data having a support (Gottschalk et al., 2011, Gottschalk et al., 2011).

In this paper, the reconstruction of missing runoff data is achieved by means of the runoff maps derivation taking into account only the spatial structural dependence and neglecting the spatial-temporal dependence. Assimilating the runoff to an areal process, a stochastic method finalized to the derivation of water balance consistent runoff maps with a geostatistical approach, has been carried out.

2. SPATIAL INTERPOLATION METHOD OF AREAL HYDROLOGICAL VARIABLE

Starting from the overview of the different methods used in literature to estimate the runoff variable, a method belonging to the geostatistical approaches has been here chosen. In particular, a modified formulation of a geostatistic method, appropriate for interpolation of an areal hydrological variable, such as runoff, has been applied. This method is based on the solution of a system of equations similar to those used for point kriging method. It takes into account both the area and the nested nature of catchments. The presented method can be seen as an interesting approach to address the problem of Prediction in Ungauged Basins (PUB) (Sivapalan et al., 2003), i.e. to estimate streamflow and streamflow-related variables at locations where no measurements are available. This method allows to estimate the runoff and so to reconstruct serially incomplete data records in basins with short streamflow records or in ungauged river basins.

The presented method is based on a disaggregation of the mean annual streamflow measured at the outlet of a basin finalized to the estimation of annual runoff on a target partition of these basins defined by the superimposition of a regular grid with a certain resolution. In this way, it is possible to obtain the estimated values of runoff in ungauged partition of certain basins or in gauged basins in which the gauging station has provided time series characterized by incomplete streamflow records.

The first step is the application of the denesting procedure at the nested basin, which allows to obtain a group of non-overlapping basins. Runoff observations might be nested i.e. the drainage basins of one station are contained in a larger basin of another station. It is therefore worthwhile to make, in a first step, a “denesting” of observed runoff within a larger basin. This disaggregation procedure can be called the “*first level of hierarchization*”.

The second step is the calculation of distances between each pair of independent drainage basins following the hierarchy of the drainage network. The appropriate distance should include the drainage network and the hierarchy of drainage basins in the system. Here, this is made possible by replacing the Euclidean distance by a "geostatistical" distance, called *Ghosh* distance. This geostatistical distance between two drainage basins is expressed as the mean of the distances between all possible pairs of points inside the two drainage basins (Gottschalk, 1993). Practically, for the group of the non-overlapping basins obtained by the application of denesting procedure, random point for each basins were produced and the average over all possible distances between pairs points in the respective catchments was calculated, the *Ghosh* distance was applied:

$$d(A_n, A_m) = \frac{1}{A_n A_m} \int_{A_n} \int_{A_m} \| \mathbf{u}_{\mathbf{A_n}} - \mathbf{u}_{\mathbf{A_m}} \| \, \mathbf{du}_{A_n} \mathbf{du}_{A_m} \qquad \forall m, n = 1, \dots, N \tag{1}$$

where $||u_{An}-u_{Am}||$ is the Euclidean distance between all random points in the areas, taken pairwise. A_n and A_m are the areas of the non-overlapping units contained in the area taken into account and N is the number of these non-overlapping basins.

Thus an empirical covariogram can be derived and, under the assumption of second-order stationarity, an empirical covariogram $Cov_e(A_n, A_m)$ is deduced. In particular, the values of covariance to represent the empirical covariogram are calculated with the following equations:

$$Cov_e(A_n, A_m) = (Q(A_n) - m_Q) * (Q(A_m) - m_Q) \qquad \forall m, n = 1, \dots, N \tag{2}$$

where N the number of areas taken into account and m_Q is the annual mean of the annual runoff of the basins. In this way, it is possible to arrange the variance and covariance matrices whose rank is equal to the number of non-overlapping basins considered.

Before making the representation of the covariogram, the main diagonal of the matrix of distance has been replaced with zeros, since it has been assumed that the distance of an area with itself must be zero whatever the criterion used to calculate it. Furthermore, at zero distance, the covariance should be equal to the variance.

At this point, an experimental covariogram can be drawn. This step is followed by a selection of possible theoretical models for the point process covariance function Cov_p. The related theoretical covariogram $Cov(A_n,A_m)$ with the local supports A_n and A_m, respectively, is derived in a similar manner by averaging the point process covariance function:

$$Cov\left(A_n, A_m\right) = \frac{1}{A_n A_m} \int_{A_n} \int_{A_m} Cov_p\left(\| \mathbf{u}_{\mathbf{A_n}} - \mathbf{u}_{\mathbf{A_m}} \|\right) \cdot \mathbf{du}_{\mathbf{A_n}} \mathbf{du}_{A_m} \tag{3}$$

The choice of the best point process model is based on a graphical comparison between *Cov* and the experimental covariogram Cov_e. Once this procedure has been carried out, it is possible to go on the ”*second level of hierarchization*”. For estimation of runoff (q) as an areal process, the following formula has been used:

$$\hat{q}(a_0) = \sum_{j=1}^{N} \lambda_j q(A_j) \tag{4}$$

where a_0 is the fundamental unit of the map (*second level of hierarchization*), A_j with $J=1,\ldots,N$ are the areas of drainage basins with observations (first level of hierarchization). Q is the column vector of observations and Λ^T is the transposed column vector of weights λ_j ($j=1,\ldots,N$), associated with the N observations.

The cornerstone is a drainage basin A_T where the mean annual discharge Q_T at the outlet point is known from measurements or estimation. In Figure 1, A_T is the total area of the considered basin and Q_T is the volume per time unit (year or month) or discharge at the outlet point, (squared yellow marker in figure) that is the sum of discharges of all sub-basins, considering the values obtained from the application of denesting procedure.

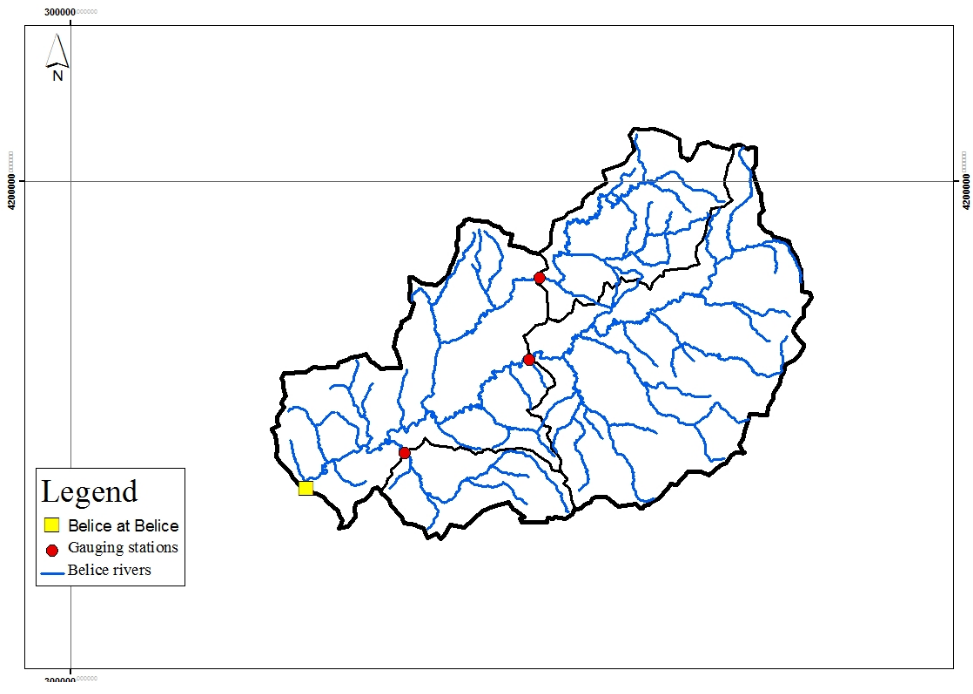

Figure 1. An example of nested basin and location of gauging stations.

The area A_T is approximated by a regular grid of n_T fundamental square cells of area a, so that $A_T=n_T*a$. It is assumed that the runoff distribution across each fundamental unit (square cells) is uniform and it is valid the following equality:

$$q_T = \frac{Q_T}{A_T} = \frac{Q_T}{n_T a} \tag{5}$$

After this preliminary stage, a second level of hierarchization has been carried out. The total area A_T has been subdivided into M non-overlapping areas ΔA_i, $(i=1,\ldots,M)$, as Figure 2 shows.

The aim is to estimate the specific discharge $q(\Delta A_i)$ for each of these areas. Such a disaggregation can naturally keep on with a stepwise disaggregation of each of the runoff $q(\Delta A_i)$ into smaller units. In the following, the algorithm for the interpolation of runoff based on the unscaled runoff $q(\Delta A_i)$ is shown.

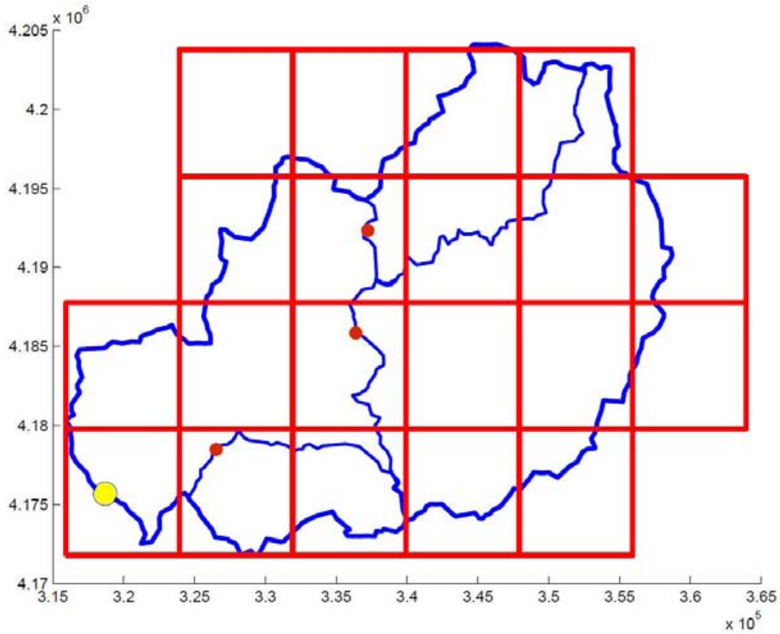

Figure 2. An example of an area A_T subdivided into M non-overlapping areas ΔA_i applied to the Belice basins (Sicily) (red dots=gauging stations, yellow dot=outlet gauging station).

Afterwards the interpolated runoff depth can be easily aggregated to the drainage basin ΔA_i, by reversing (5) to assess discharge values:

$$Q(\Delta A_i) = n_i a q(\Delta A_i) \tag{6}$$

where n_i is the area of the unit ΔA_i, measured in terms of number of cells.

If discharge observations are available at N basins with area A_j $j=1,\ldots,N$ as above, the insertion into (4) yields the following equation for interpolation of the specific runoff:

$$q(\Delta A_i) = \sum_{j=1}^{N} \lambda_j^i q(A_j) = \Lambda^T H \tag{7}$$

The optimal weights in (7) (or 4) are found by minimizing the estimation variance. Assuming a local second-order stationarity of the process and under the condition of unbiasedness this leads to the following linear equation system for the calculation of weights λ_i^j $(j=1,\ldots,N)$

$$\Lambda_j = C^{-1} C_{0j} \tag{8}$$

with constraint

$$\sum_{i=1}^{M} \lambda_i^j = 1 \tag{9}$$

with

$$\Lambda_i = \begin{bmatrix} \lambda_1^i \\ \lambda_2^i \\ \vdots \\ \lambda_N^i \\ \mu^i \end{bmatrix} \qquad C = \begin{bmatrix} Cov(A_1, A_1) & Cov(A_1, A_2) & \dots & Cov(A_1, A_N) & 1 \\ Cov(A_2, A_1) & Cov(A_2, A_2) & \dots & Cov(A_2, A_N) & 1 \\ \dots & \dots & \dots & \dots & \dots \\ Cov(A_N, A_1) & Cov(A_N, A_2) & \dots & Cov(A_N, A_N) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \qquad C_0 = \begin{bmatrix} Cov(A_1, \Delta A_1) \\ Cov(A_1, \Delta A_2) \\ \vdots \\ Cov(A_1, \Delta A_N) \\ 1 \end{bmatrix}$$

The only constraint to the weights is the total sum equal to one, and this does not exclude the presence of negative values. The streamflow at the outlet point of basin A_T is given by:

$$\sum_{i=1}^{M} \Delta A_i q(\Delta A_i) = \sum_{i=1}^{M} n_i a q(\Delta A_i) = \sum_{i=1}^{M} n_i a \left(\sum_{j=1}^{N} \lambda_j^i q(A_j) \right) \tag{10}$$

The sum of the interpolated discharge for each of the sub-basins, calculated from (10), does not necessarily match the discharge Q_T observed downstream. A further step is to include a constraint so that the interpolated lateral inflow is balanced with the observed runoff in the river system. Rearrangement of (10) gives

$$\sum_{i=1}^{M} n_i a \left(\sum_{j=1}^{N} \lambda_j^i q(A_j) \right) \Rightarrow \sum_{i=1}^{M} \left(\sum_{j=1}^{N} n_i \lambda_j^i q(A_j) \right) = n_T q_T \tag{11}$$

This new constraint integrates the previous ones presented in (9). The (7) remains the same for this case, but the weights λ^i_j (j=1,…,N, i=1,…,M) have to be calculated simultaneously for all M elements. Optimal weights were found through the solution of the system of equations:

$$\Lambda = C^{-1} C_0 \tag{12}$$

with

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \vdots \\ \mathbf{L}_M \\ \mu_T \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} \mathbf{K} & 0 & \cdots & 0 & \mathbf{V}_1 \\ 0 & \mathbf{K} & 0 & \dots & \mathbf{V}_2 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \dots & 0 & \mathbf{K} & \mathbf{V}_M \\ \mathbf{V}_1^T & \mathbf{V}_2^T & \dots & \mathbf{V}_\mathbf{M}^\mathbf{T} & 0 \end{bmatrix} \quad \mathbf{C_0} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_M \\ n_T q_T \end{bmatrix} \quad L_i = \begin{bmatrix} \lambda_1^i \\ \lambda_2^i \\ \vdots \\ \lambda_N^i \\ \mu^i \end{bmatrix} \quad K = \begin{bmatrix} Var(A_1) & Cov(A_1, A_2) & \dots & Cov(A_1, A_N) & 1 \\ Cov(A_2, A_1) & Var(A_2) & \dots & Cov(A_2, A_N) & 1 \\ \dots & \dots & \dots & \dots & \dots \\ Cov(A_n, A_1) & Cov(A_n, A_2) & \dots & Var(A_N) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad \mathbf{V_i} = \begin{bmatrix} n_i q(A_1) \\ n_i q(A_2) \\ \vdots \\ n_i q(A_N) \\ 0 \end{bmatrix} \quad \mathbf{G}_i = \begin{bmatrix} Cov(A_1, \Delta A_1) \\ Cov(A_1, \Delta A_2) \\ \vdots \\ Cov(A_1, \Delta A_N) \\ \mu^i \end{bmatrix}$$

Then, using the (7) it is possible to obtain an estimate of the runoff in the areas ΔA_i.

3. CASE STUDY

The dataset used in this study come from 105 hydrometric stations distributed throughout Sicily.

In order to obtain a reliable database of runoff, in the process of data retrieval, the monthly and annual streamflows recorded in the period ranging from 1923 to 2002, for a total of 80 years observed, have been minutely examined. In order to ensure greater reliability for further analysis, the gauge stations that have worked for less than 10 years have been removed; so the initially available hydrographic information has reduced from 105 to 69 stations.

The methodology is applied to 23 Sicilian main basins containing 58 sub-basins.

For the application of this method, the mean annual runoff of each basin, the area of basins and the working period of the gauging station are taken in account. Not all the selected station worked for the same period. So, in order to consider the period of years in which the hydrometric stations have operated less discontinuously as possible, a limited time window is taken into account (from 1960 to 2002).

Moreover, because of the heterogeneity that characterizes the climate and morphology in Sicily (the total annual precipitation varies between 400 and 1200 mm), a subdivision of the Sicily region has been made. In particular, this analysis has been performed dividing the island into three zones (Figure 3), using the homogeneous regions suggested by Cannarozzo et al. (1995):

1. Zone 1: the most of the catchments (32) belongs to the Zone 1, which is the Northwestern part of the island where the mean annual rainfall is around 680 mm, close to the regional value. The average area of the examined basins in this area is 200 Km^2, ranging from 10 up to 1186 Km^2.
2. Zone 2: this Zone 2 has the lower number of stations (12), but it is also the smallest sub-area. The mean annual rainfall is around 900 mm, higher than the regional value and the basins inside this zone are characterized by relatively small size and steep slopes, especially in the Northeastern part.
3. Zone 3: this Zone 3 is located in the South- East part of the island and contains 14 stations. The average annual rainfall, equal to 620 mm, is lower than the regional value and the average size of the considered basins is about 300 Km^2.

The homogeneity of these regions has been tested in terms of annual streamflow (Cannarozzo et al., 2009) using the homogeneity test of Hosking and Wallis (1997).

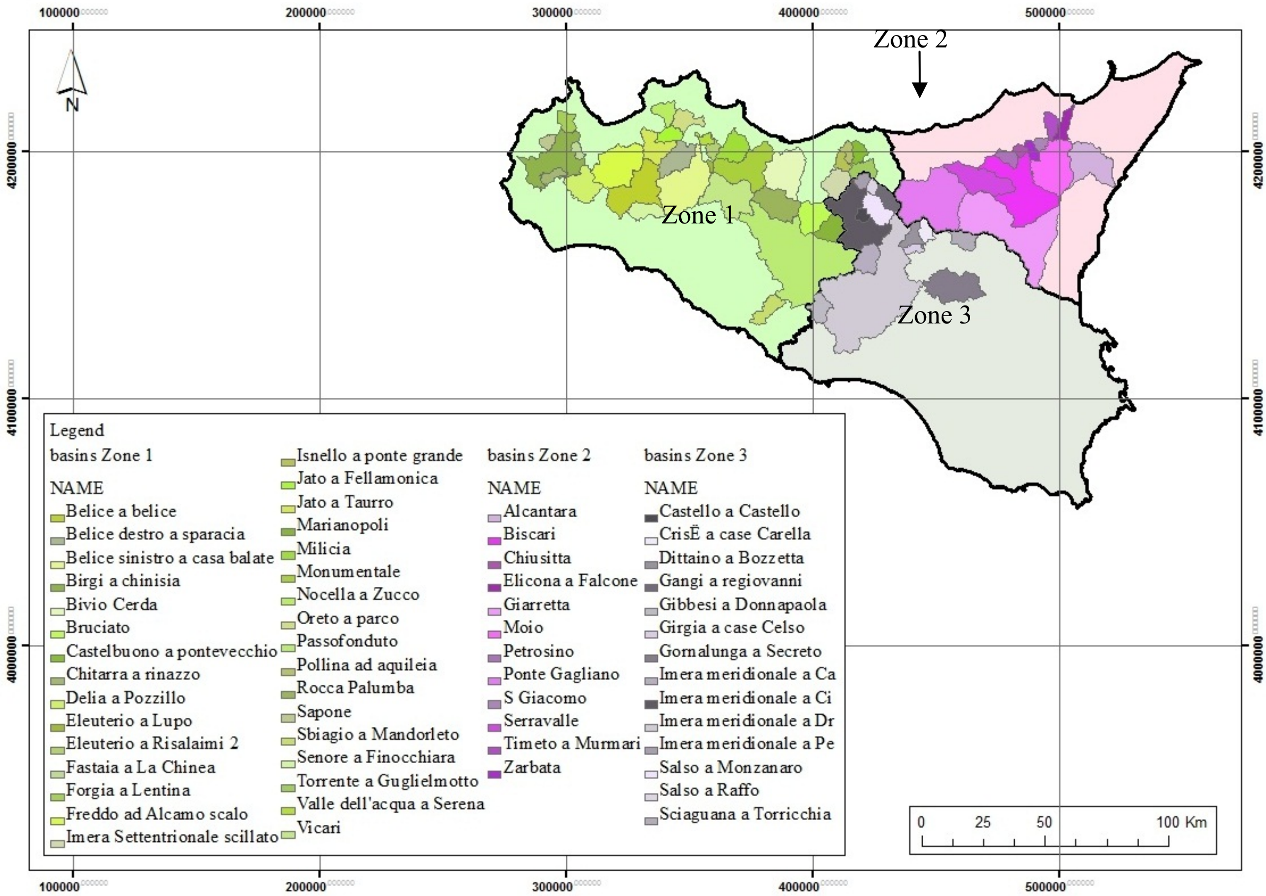

Figure 3. Sicily region subdivided in three zones and overlapping of catchments.

4. APPLICATION OF THE MAPPING PROCEDURE

The applied methodology is based on subsequent levels of hierarchization. The number of levels in this hierarchy is determined mainly by the amount of available observations, which also indicates the level of detail that can be achieved in terms of size and number of fundamental units of the map. As mentioned above, the first level in a larger drainage basin is usually already well defined by existing observation stations in the main rivers constituting the first level of sub-basins (*first level of hierarchization*). These basins are, in turn, divided into a second level of sub-basins (or grid cells) (*second level of hierarchization*), and observation stations with appropriate basin scales are chosen as the background for the interpolation. The ”*first level of hierarchization*” is the denesting or disaggregation procedure. In this case, the method is applied to the nested basins belonging to the three zones above defined (*Zone 1*, *Zone 2* and *Zone 3*). From the application of disaggregation procedure, *N* sub-basins are obtained for each zone, within the each considered zone (*N=32* in *Zone 1*; *N=12* in *Zone 2*; *N=14* in *Zone 3*). The value of runoff for each of different non-overlapping basins was obtained taking into account the network structure of the basin. For sake of simplicity, the different steps of mapping procedure are graphically shown only for zone 1. In Figure 4, the average annual runoff estimated by the disaggregation procedure for different areas is shown.

Applying the disaggregation method to all Sicilian basins, the presence of negative runoff values has been observed in such basins. Such circumstance clearly shows an error in data capture and data processing by UIR (*Ufficio Idrografico Regionale*, now known as OA-ARRA - *Osservatorio delle Acque - Agenzia Regionale dei Rifiuti e delle Acque*).

Because of this, the data are not reliable and sometimes not useful for hydrological modeling. In this case, when the runoff values, obtained by the methods previously explained, are negative, they are removed from dataset. Then, the calculation of the *Ghosh* distance and the theoretical and empirical covariograms has been made. In practice, for the *Ghosh* distance calculation, 100 random points for each basin are produced and the Euclidean distance between all random points in the area, taken pairwise, is made.

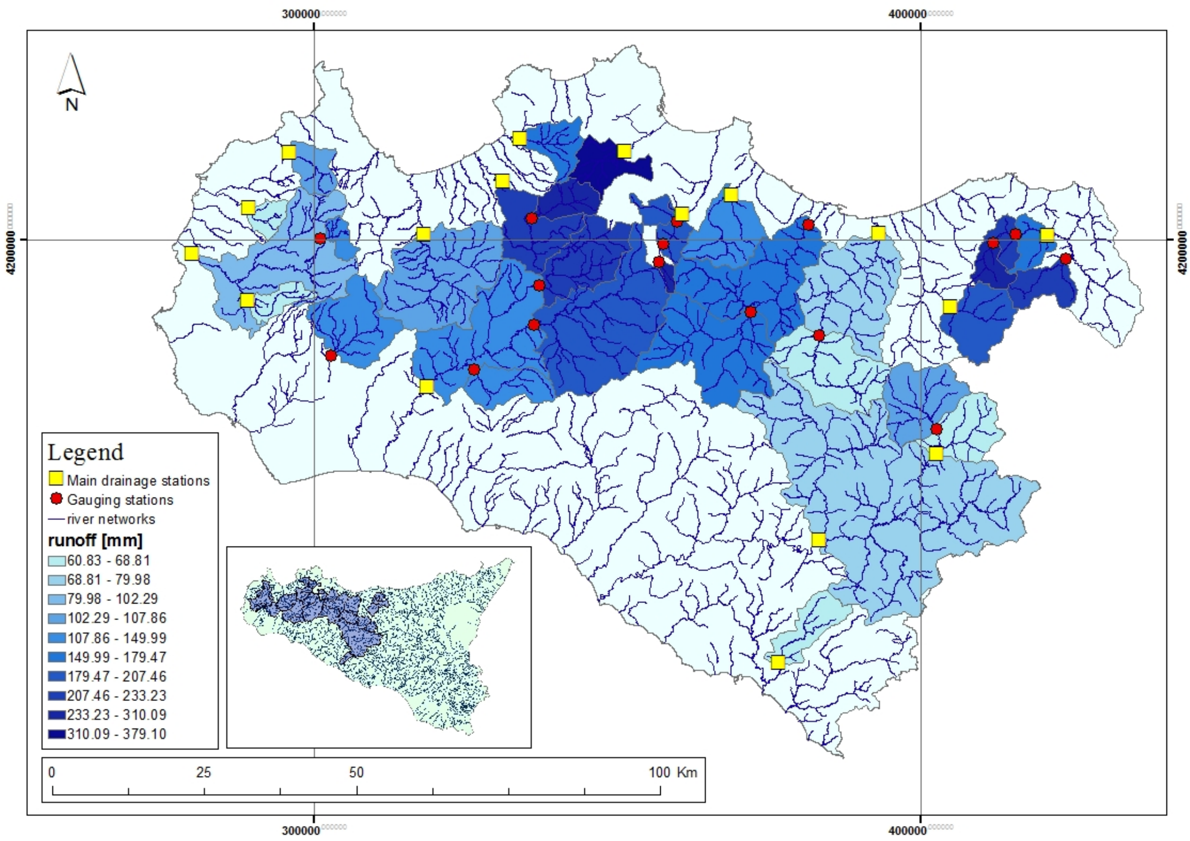

Figure 4. The average annual runoff estimated by the disaggregation procedure for the Zone 1.

With regarding to the choice of the best theoretical covariogram model, in order to fit the experimental covariogram, it is necessary to carry out some considerations. For the measured runoff data used in this study, a singular spatial correlation structure has been observed. A non-parametric equation of the covariogram function (Ploner and Dutter, 2000) is here suggested:

$$Cov(d) = C(0) * \left[\left(\left(1 - \frac{d}{R} \right) * exp\left(-\frac{d}{R} \right) \right) \right] \tag{13}$$

This model function can be fitted by minimizing the target function F for the parameter R:

$$F(R) = \left(Cov(d) - C(0) * \left[\left(\left(1 - \frac{d}{R} \right) * exp\left(-\frac{d}{R} \right) \right) \right] \right)^2 \tag{14}$$

where *C(0)* is the covariance value at zero distance.

The theoretical covariogram is represented with a lag equal to 8000 m (Figure 5). The spatial scale coefficient was set to $R=16882.24m$ and $Cov_e(0)=2.894*10^{14}m^6$ in the *Zone 1*, $R=21176.373m$ and $Cov_e(0)=2.363*10^{15}m^6$ in the *Zone 2*, $R=10589.027m$ and $Cov_e(0)=5.871*10^{14}m^6$ in the *Zone 3*.

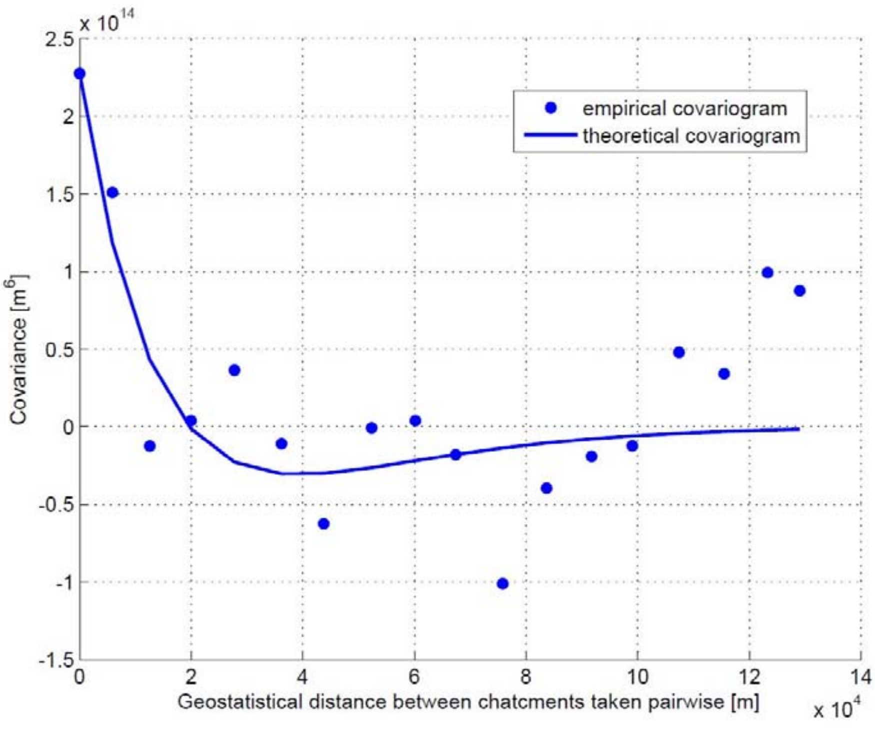

Figure 5. Experimental and theoretical covariograms: Zone 1.

In order to illustrate the principle of the disaggregation procedure, the interpolation scheme is applied to a target partition defined by the superimposition of a regular 8 x 8 km grid (64 km^2) over the catchments boundaries belonging to the three different zones taken into account ("*second level of hierarchization*") (Figure 6). In this case, the point of departure is given from more than one drainage basin A_T. Now, for each zone, the areas A_{Tk} (with $k=1,...,m_d$, where m_d is the number of outlet sections of the major drainage basins) are taking into account. The procedure is the same that described in section 2. In particular, now, a constraint has been considered for all major drainage basins in the examined area and k systems of equations 14 will be applied (one for each constraint). The interpolation procedure to assess runoff on 2 x 2 km cells (4 km^2) ("*third level of hierarchization*") is applied to the full data set and the interpolation constraint is kept within each 8 x 8 km cell so that the sum of runoff volume from the smaller cells equals the runoff volume from this larger one. The sum of runoff volume from the larger cells is, in its turn, equals the runoff volume from the total drainage basins, i.e. the sum of the runoff volume of the basins belonging to the zone taken into account. The expected pattern of runoff structure has been reproduced on the six maps (two for each zone), shown in Figures 7 and 8.

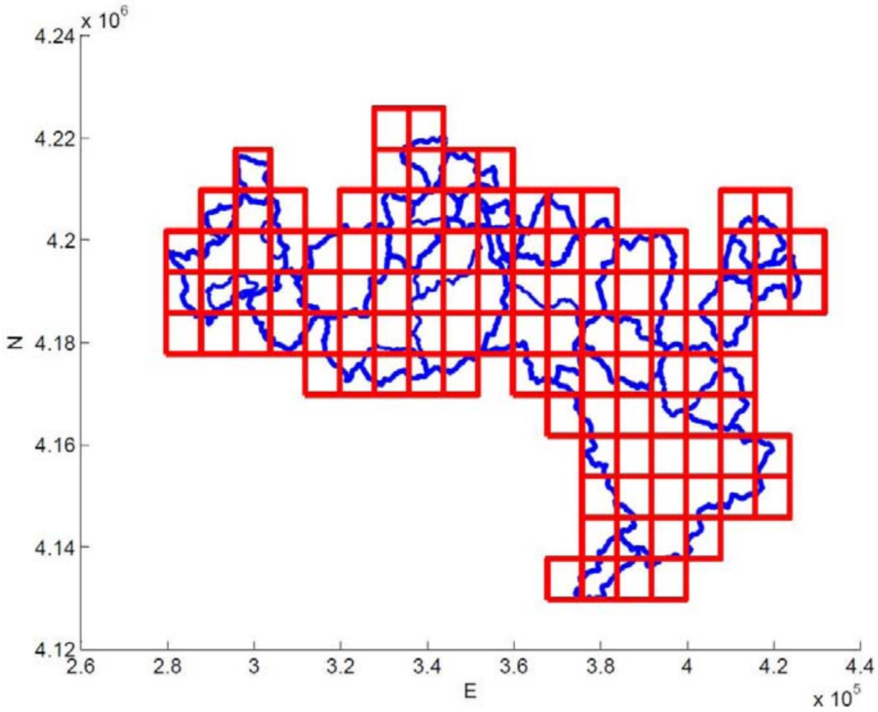

Figure 6. Total area A_T (basins within Zone 1) subdivided into M non-overlapping areas ΔA_i.

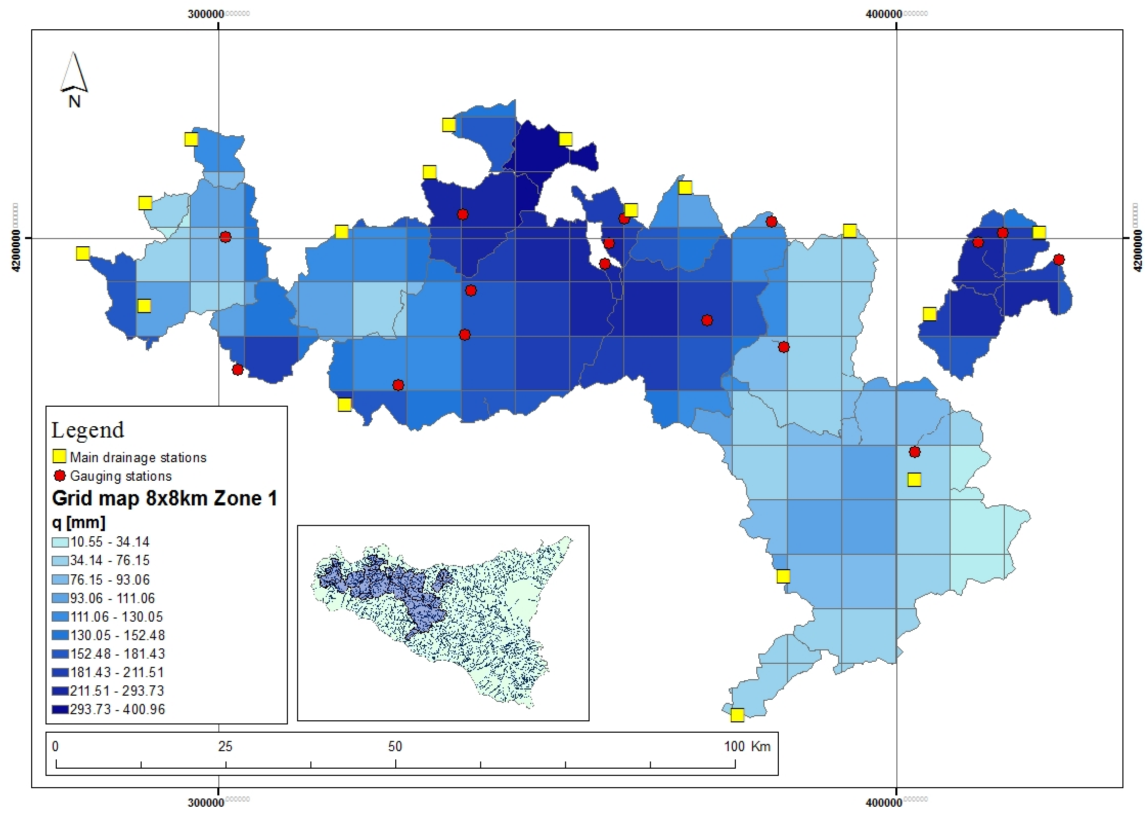

Figure 7. Gridded map of average annual runoff with 8 x 8 km resolution (Zone 1).

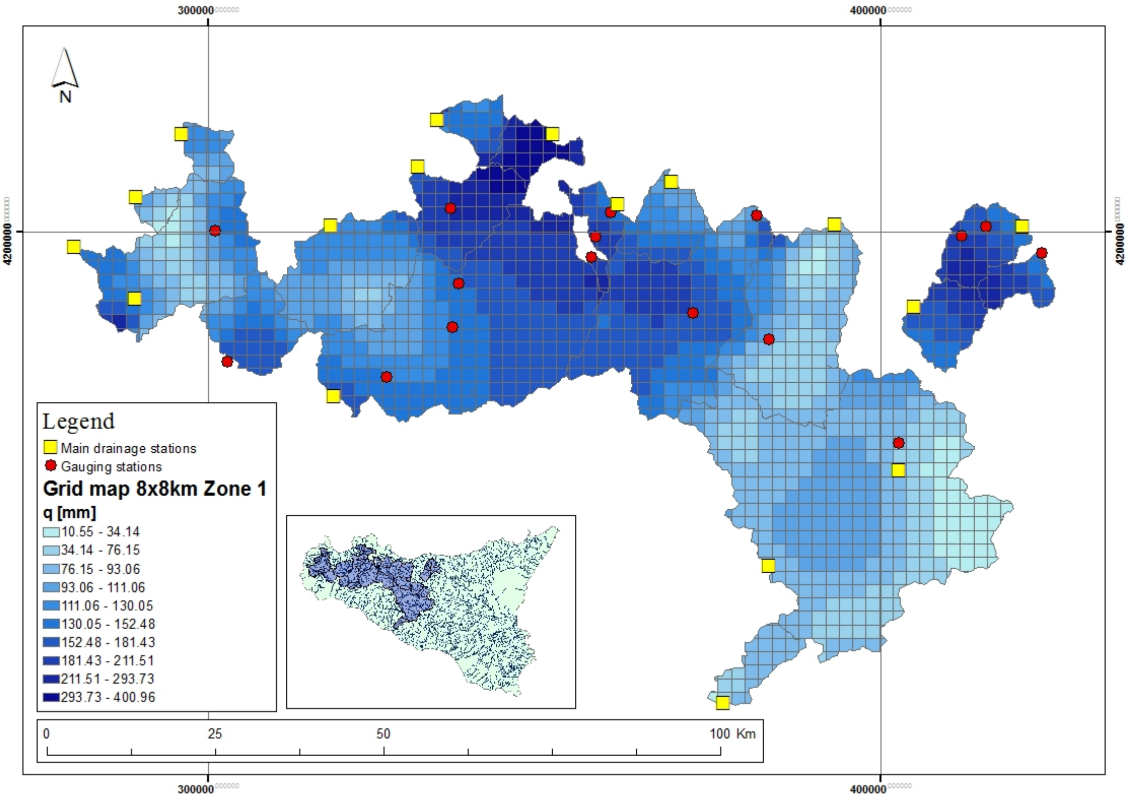

Figure 8. Gridded map of average annual runoff with 2 x 2 km resolution (Zone 1).

4.1 Validation

A cross validation is performed for this mapping technique, in order to validate the applied methodology. This analysis consists of excluding one gauging station in turn from the group of basins and then estimating runoff at this site by applying the interpolation procedure to the remaining stations. The cross-validation analysis provides valuable information about the real influence of the global constraint on runoff assessment. For each removed station, the observed runoff is compared to the estimated runoff. This analysis is applied, for each zone, to the second level of hierarchization.

The performances of the interpolation method have been assessed using two indexes, taking into account the estimated values of runoff obtained with the validation procedure. In particular, the correlation coefficient (*CC*) has been calculated together with the relative deviation (*RD*) used to assess the percentage difference between estimated and observed mean annual runoff:

$$RD = \frac{(q_{obs} - q_{est})}{q_{obs}} * 100 \quad (\%) \tag{15}$$

where, q_{obs} and q_{est} are respectively observed and estimated mean annual runoff values.

The performance of the method for the *Zone 1* (Figure 9) are quite good and the *CC* is equal to 0.99. The algorithm provides not very good results when the stations with the lowest values of observed runoff are removed. In this case, an overestimation of the observed runoff values can be seen when *Baiata at Sapone* and *Chitarra at Rinazzo* are removed from the data set. These sub-basins are characterized by the lowest runoff values in *Zone 1*. The highest *RD* is observed for the *Chitarra at Rinazzo* sub-basin and is equal to 21%. Another high value of RD is obtained for the estimated value of runoff when the *Fastaia at La Chinea* is removed from the data set. In particular *RD*, in this case, is equal to 18%, suggesting an underestimation of runoff value.

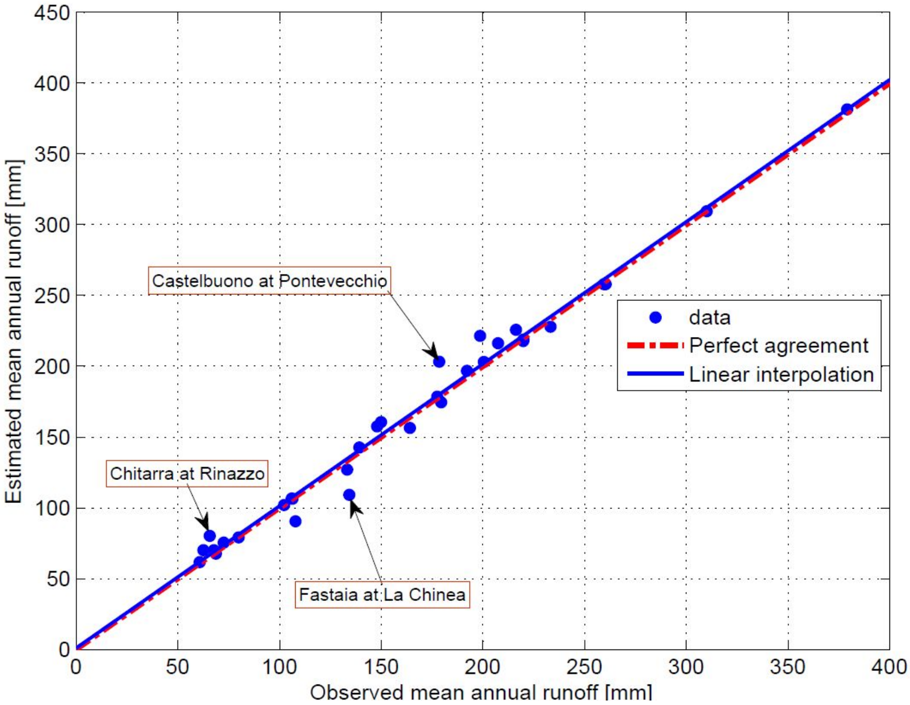

Figure 9. Cross-validation of annual runoff (Zone 1).

Similar performances can be observed in *Zone 2* where the correlation coefficient is equal to 0.97. In this case, the values of runoff of *Saraceno at Chiusitta* and *Alcantara at San Giacomo* were underestimated. The highest *RD* of estimate is for the *Saraceno at Chiusitta* sub-basin and is equal to 23%, while the *RD* is equal to 13% when *Alcantara at San Giacomo* is removed.

With regard to the performance of method for the *Zone 3*, it is possible to observe that the results are quite worse than the results of the first two zones, even if *CC* is equal to 0.96. In this case, an underestimate of the observed runoff values is noted when *Imera Meridionale at Petralia* sub-basin is removed from the dataset (*RD* is equal to 23%). Another high value of *RD* is obtained for the estimate value of runoff when the *Imera Meridionale at Capodarso* is removed from the data set (*RD*=24%), suggesting an underestimation of runoff value. Moreover, an overestimation of observed runoff is obtained for *Gangi at Regiovanni* and *Gibbesi at Donnapaola*. In these cases the highest values of *RD* has been achieved: *RD* equal to 36% for *Gangi at Regiovanni* and *RD* equal to 60% for *Gibbesi at Donnapaola*.

In order to obtain estimated runoff values also in other region not belonging to the considered basins, the interpolation scheme is applied to the total area of the considered zones. In each of these regions, not belonging to the considered basins, the same procedure, above described, is applied to the basins belonging to the considered zone using the same grid size.

The use of Digital Elevation Model (DEM) of Sicily and spatial analysis techniques in a GIS environment (ESRI ArcGIS) have allowed the derivation of streamflow discharges over the Sicily (Figure 10). In particular, the function *WeightedFlowAccumulation* has provided the derivation of a grid containing the streamflow discharges using as weight grid the spatial distribution of runoff.

5. CONCLUSIONS

For the runoff variable, estimated by a mapping technique, a cross-validation is performed to test the obtained results. It is important to highlight that in this case only an annual scale analysis has been performed to obtain the gridded maps of estimated average annual runoff.

The cross-validation analysis gives valuable information about the real influence of the global constraint on runoff in the application of the method. On the contrary, the accuracy of the method on sub-basins partition (gridded maps of average annual runoff with 8 x 8 km and 2 x2 km resolution) can be only assessed verifying if the partition strictly respects or not the hierarchical structure of the catchment in comparison with the observed runoff pattern. Indeed, no objective validation can be proposed because of the lack of reliable measurement and the maps are analyzed on visual agreement with observed runoff patterns. So, taking into account the information obtained by the cross validation methods, for all groups of basins in the considered zones *(Zone 1* - western area of Sicily, *Zone 2* - eastern of Sicily and in *Zone 3* - south-eastern of Sicily), the results are quite good, in terms of real influence of the global constraints. The algorithm gives poor results when the observation removed from the dataset is one of the extreme values (highest or lowest). From the visualization of the maps, it is possible to observe that in *Zone 1* and *Zone 2* a good agreement between the observed and estimated runoff patterns. Furthermore, almost all the partitions strictly respect the hierarchical structure of the catchment. A different situation is encountered for *Zone 3*, where the presence of negative runoff estimated values demonstrates that the methods does not succeed in reproducing correctly the runoff pattern and most of the partitions do not strictly respect the hierarchical structure of the catchment. This is probably due to a low quality of the input data.

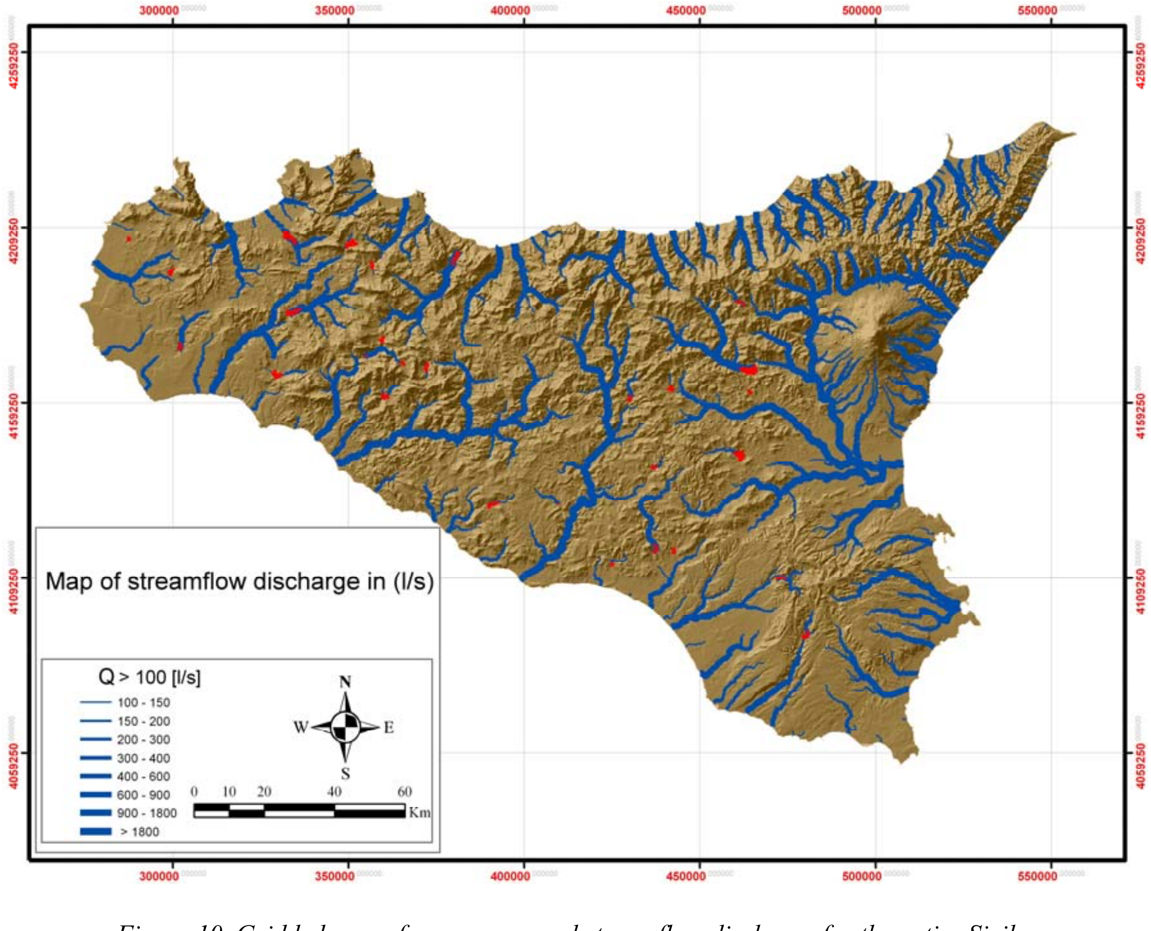

Figure 10. Gridded map of average annual streamflow discharge for the entire Sicily.

Finally, the application of this method gives the annual runoff estimated data for the stations that have been out of work in the chosen time window and that are characterized by a dataset affected by missing data. Moreover, since the hierarchical principle allows the calculation of gridded maps for finer and finer resolution annual runoff estimated values can be obtained also for the areas of the basin not provided with gauge stations.

ACKNOWLEDGEMENTS

We would like to thank Prof. Lars Gottschalk (University of Oslo) for his thoughtful suggestions and helpful comments.

REFERENCES

Arnell, N. W., 1995. Catchments as space time filters; a joint spatio temporal geostatistical analysis of runoff and precipitation. Journal of Hydrology; 167: 39-56.

Cannarozzo, M., D'Asaro, F., and Ferro, V., 1995. Regional rainfall and flood frequency-analysis for Sicily using the 2-component extreme-value distribution. Hydrological Science; 40: 19-42.

Cannarozzo, M., Noto, L.V., Viola, F., La Loggia, G., 2009. Annual runoff regional frequency analysis in sicily, physics and chemistry of the earth, Physics and Chemistry of the Earth, 34, 679-687.

Cutore, P., Cristaudo, G., Campisano, A., Modica, C., Cancelliere, A., Rossi, G., 2006. Regional Models for the Estimation of Streamflow Series in Ungauged Basins. Water Resources Management; 21(5): 789-800.

Gottschalk, L., 1993. Correlation and covariance of runoff. Stochastic Hydrolog. Hydraul.; 7: 85-101.

Gottschalk, L., Krasovskaia, I., 1998. Development of Grid-related Estimates of Hydrological Variables. Report of the WCP-Water Project B.3, WCP/WCA, Geneva, Switzerland.

Gottschalk, L., Krasovskaia, I., Leblois, E., Sauquet, E., 2006. Mapping mean and variance of runoff in a river basin. Hydrology and Earth System Sciences Discussions; 3(2): 299-333.

Gottschalk, L., Leblois, E., Skøien, J., 2011. Correlation and covariance of runoff revisited. Journal of Hydrology, doi:10.1016/j.jhydrol.2010.12.011.

Gottschalk, L., Leblois, E., Skøien, J., 2011. Distance measures for hydrological data having a support. Journal of Hydrology. doi:10.1016/j.jhydrol.2011.03.020.

Hosking, J., Wallis, J., 1997. Regional frequency analysis: An approach based on l-moments, Cambridge Univ. Press, New York.

Ilunga, M., Stephenson, D., 2005. Infilling streamflow data using feed-forward backpropagation artificial neural networks: Application of standard BP and pseudo Mac Laurin power series BP techniques. WaterSA; 31(2): 171-176.

Maskey, S., 2009. Modelling uncertainty in flood forecasting system. London, UK: *Taylor and Francis e Library.*

Ploner, A., Dutter, R., 2000. New directions in geostatistics. Journal of Statistical Planning and Inference; 91(2): 499-509.

Sauquet, E., Gottschalk, L., Leblois, E., 2000. Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme. Hydrological Sciences Journal; 45(6): 799-815.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. Iahas decade on predictions in ungauged basins pub, 2003:2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences; 48: 857-880.

Skøien, J. O., Bloschl, G., 2005. Top kriging; geostatistics on stream networks. Hydrology and Earth System Sciences Discussions; 2(6): 2253-2286.

Skøien, J. O., Bloschl, G., 2006. Catchments as space time filters; a joint spatio-temporal geostatistical analysis of runoff and precipitation. Hydrology and Earth System Sciences; 3(3): 941-985.