To be cited as: Chebbi , A., Bargaoui , Z.K., Abid, N., Cunha, M. C., "Optimization of a hydrometric network extension using specific flow, kriging and simulated annealing" Journal of Hydrology, 555, pp. 971-982, 2017. doi: 10.1016/j.jhydrol.2017.10.076

Optimization of a hydrometric network extension using specific flow, kriging and simulated annealing

Afef Chebbi[1,2] , Zoubeida Kebaili Bargaoui[1] , Nesrine Abid[1] and Maria da Conceição Cunha[3]

[1] Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, BP 37 1002 Tunis, Tunisia

[2] Université de Sousse, Institut Supérieur des Sciences Appliquées et Technologie de Sousse, Cité Taffala (Ibn Khaldoun), 4003 Sousse, Tunisia

[3] Civil Engineering Department, University of Coimbra Polo II da Universidade-Pinhal de Marrocos, 3030-290 Coimbra, Portugal

Abstract

In hydrometric stations, water levels are continuously observed and discharge rating curves are constantly updated to achieve accurate river levels and discharge observations. An adequate spatial distribution of hydrological gauging stations presents a lot of interest in linkage with the river regime characterization, water infrastructures design, water resources management and ecological survey. Due to the increase of riverside population and the associated flood risk, hydrological networks constantly need to be developed. This paper suggests taking advantage of kriging approaches to improve the design of a hydrometric network. The context deals with the application of an optimization approach using ordinary kriging and simulated annealing (SA) in order to identify the best locations to install new hydrometric gauges. The task at hand is to extend an existing hydrometric network in order to estimate, at ungauged sites, the average specific annual discharge which is a key basin descriptor. This methodology is developed for the hydrometric network of the transboundary Medjerda River in the North of Tunisia. A Geographic Information System (GIS) is adopted to delineate basin limits and centroids. The latter are adopted to assign the location of basins in kriging development. Scenarios where the size of an existing 12 stations network is alternatively increased by 1, 2, 3, 4 and 5 new station(s) are investigated using geo-regression and minimization of the variance of kriging errors. The analysis of the optimized locations

from a scenario to another shows a perfect conformity with respect to the location of the new sites. The new locations insure a better spatial coverage of the study area as seen with the increase of both the average and the maximum of inter-station distances after optimization. The optimization procedure selects the basins that insure the shifting of the mean drainage area towards higher specific discharges.

Keywords hydrological gauging stations; network optimization; geo-regression; ordinary kriging; simulated annealing

1. Introduction

A hydrometric network is aimed at giving the hydrological information to be used for ecological survey, hydrological survey, hydrological regionalization as well as infrastructures design. Flood estimates are of major importance since they are needed for designing civil engineering works, inundation risk zoning and an estimation of ecological flows. Both water source infrastructure design and management (reservoirs, water distribution systems, irrigation networks, etc.) are based on flood estimation. Due to the increase of riverside population and the associated flood risk issues, the hydrological networks need to be developed.

According to Mishra and Coulibaly (2009), a hydrometric network should be optimized to collect most hydrological information and in the most precise way. More generally, the commonly used processes for network optimization include statistical approaches, a user survey procedure, a hybrid approach, and sampling plans (Vivekanandan, 2012). *Statistical approaches* for hydrometric network optimization range from clustering methods (Bum and Goulter, 1991) and spatial regression (Tasker and Stedinger, 1989) to entropy-based techniques (Caselton and Husain, 1980). *Clustering methods* are usually used to identify groups of hydrometric gauging stations with similar flow characteristics on the basis of a similarity matrix defining the similarity of each station to every other station. This constitutes an important step in the network design procedure. The annual average runoff is a main flow characteristic and spatial regression is often used to predict it at ungauged locations (Daigle et al., 2011). *Entropy methods* may also assist network design by quantifying the relative information content and by estimating incertitude (Vivekanandan, 2014). Moreover, the *User survey procedure* is based on the users' needs to continue or discontinue stations depending upon the type of data needed in the basin. This investigation by its nature relies on a certain amount of personal decisions (Davar and Brimley, 1990).

The hybrid method combines models by adopting the output from one method as an input into another model for network optimization. For example an algorithm of numerical optimization permits to improve the optimal network design by variance reduction and allows the insertion of other criteria in the objective function such as the economic cost of the data collection (Mishra and Coulibaly, 2009). *Hydrologic sampling* plans are based on the influence of rainfall on stream flow processes. The effectiveness of sampling plans is evaluated by the variance of error in the estimate stream flow (Tarboton et al., 1987).

On the other hand, the rainfall network design is often achieved by using the kriging interpolation method combined with optimization algorithms such as simulated annealing (see for example Barca et al., 2008; Chebbi et al., 2013). Kriging has also been used for piezometric networks optimization. For instance, Rouhani (1985) used two criteria for piezometric network optimization: the first concerns the reduction of the kriging variance while the second is related to the expected economic gain, measured by loss reduction. One fundamental upshot of kriging is that it results in the estimation of the variance of interpolation errors, making it possible to evaluate network performance. Whereas entropy method is worth for existing networks, the kriging interpolation method may be extended for planned networks. Kriging often employs a semivariogram function representing the structure of the spatial variability of the data. The semivariogram effectively gives the same information as an auto-correlation function. However, it has a big advantage of being an unbiased estimator as it does not depend on the mean of the data set. So, it is proposed here to get profit of the kriging approaches in order to improve the design of a given hydrometric network. The main difficulty here resides in defining a suitable hydrometric study variable and a suitable objective function, as well in addition to a suitable kriging method.

In this study, we have adopted a specific discharge as a prime study variable representing the ratio of the river discharge to the drainage area and which is also called average specific annual module. For a long time in flood studies, the record specific discharges are adopted as a key variable to obtain regionally-developed curves (Castellarin, 2007). So, a specific discharge is considered here as a key watershed descriptor.

There are many other ways to handle the issue of hydrometric network optimization since the hydrologic response is multidimensional. Therefore, instantaneous hydrograph responses to rainfall events are described by at least three variables: flood duration, flood peak and flood volume. An objective function including these variables may be achieved but we cannot rely on this approach because of data limitations. We have no information about the flood series (except at daily resolutions). Basins have signatures which can be described by using some statistics of the basis of the flow-duration curve (Sadegh et al., 2016) obtained by analyzing daily discharges. These above-mentioned statistics may be used to optimize the hydrometric network. The only statistics adopted here is the sample mean of annual discharges. We did not apply other statistics even though they would be a possible extension of the current work. The Runoff coefficient is another basin signature which can be adopted to solve the optimization of hydrological networks. The difficulty with basin runoff coefficient is that it involves the estimation of the basin average rainfall, which in turn is a “rainfall product” that needs interpolation tools in order to be evaluated. Another alternative is the use of digital models (based on a Geographic Information System) associated to soil, land use information and classification methods to find the most representative basins. The advantage of not adopting such an alternative is to limit the need of implementing digital models which themselves are to be verified using in situ data.

Thus, this work intends to extend the use of a specific river discharge, as a study variable to the hydrometric network optimization. One implicit assumption is that the geographic regions in the study are hydrologically homogeneous.

Many basin attributes may be included as a proxy for flood (and the specific discharge) estimation. They are often adopted in geo-regression approaches. The drainage area, the basin geology together with land use descriptors, soil characteristics, elevation data, and climate variables such as mean annual precipitation are often proposed as flood proxy or surrogates (Acreman and Sinclair, 1986). Wilson and Gallant (2000) noticed that steepness can be considered as a surrogate for overland and subsurface flow velocity and the runoff rate. Hundecha and Bardossy (2004) adopted basin size, slope and shape as characteristics for regionalizing Hydrologiska Byrans Vattenbalansavdelning (HBV) rainfall runoff model parameters. Kjeldsen and Jones (2007) adopted both the drainage area and the average annual rainfall together with an index of flood reduction attributable to reservoirs and lakes and a derived base flow index using the Hydrology of Soil Type classification.

Here, the drainage area, which is the most commonly used variable in the literature, is adopted as a proxy variable for the estimation of the specific module, similarly to Kron and Willems (2002) who consider only the drainage area as proxy for flood discharge for a large-scale flood hazard mapping. However, another alternative linking basin runoff to mean basin precipitation is tested. The ordinary kriging estimation involves the basin centroid inter-distances. Topological kriging (or top-kriging) is recently proposed as an alternative to ordinary kriging. It is based on regularized semivariograms between catchments which are estimated on the basis of point semivariograms and the distances between basin centroids and drainage areas are assumed as a proxy. The main difference is that top-kriging takes into account the nested nature of catchments by considering that the area is shared by two catchments. Yet, top-kriging requires a very large computation time compared to ordinary kriging. Laaha et al. (2014) found that for locations without upstream data points, the performances of the two methods are similar. Their study resulted in coefficients of determination in cross-validation that are 0.75 for the top-kriging and 0.68 for regional regression methods, including nested basins. A major interest of the top-kriging method is its ability to estimate (and allow to visualize) continuously the spatial variability of the specific flow over the whole hydrographic network. Nevertheless, in this study, we do not need to continuously estimate the specific flow rate. Therefore, in our opinion, the small gain in terms of explanatory power does not justify such an investment in computation time, especially that the kriging procedure is repeated as many times as it is necessary to optimize the objective function.

Thus, the approach using ordinary kriging is selected as an alternative. It is also achieved in order to take advantage of the numerical tools developed so far by the authors in previous studies (Chebbi et al., 2011).

To assign a geographical distance between basins (in semivariogram analysis and kriging estimation), the Euclidian distance between the basin's centroids is often adopted (see for example Daviau et al., 2000; Adamowski and Bocci, 2001; Eaton et al., 2002; Skøien et al., 2003). In fact, it is not possible to consider the basin outlets for distance estimation because the runoff is a response of the basin as a whole. Some variables other than the geographic location by such as a basin mean altitude, basin slope, and basin mean annual precipitation can be adopted to build the distances between basins but for the reasons advocated above (lack of data availability), this is out of the scope of this study.

The main purpose of this study is to identify an optimal set of new locations to upgrade the size of an initial hydrometric network. The objective addressed in stating the optimization problem is to make a more accurate evaluation of the average specific annual module.

The new contribution of this study is really to find a substitution variable for the runoff which is not suitable for the use of kriging because it is not an additive variable. The problem is solved by using the transformation of the runoff into an effective rainfall (by using the ratio of runoff and the drainage area which corresponds to the specific discharge) and also by using a scaling formula (geo-regression) allowing a basin runoff inter-comparison.

The case study concerns the hydrometric network of the transboundary Medjerda River, in Northern Tunisia. This study area is selected because the Medjerda represents the main river in Tunisia with a 350 km length. The drainage area of the basin at the Mediterranean outlet in Kallat Landlous is about 23 500 km². Another reason for which this study area is chosen relies in taking advantage of the long series of runoff observations available in this basin for a long time (Rodier et al., 1981).This insures a good accuracy in the estimation of the mean annual runoff.

Section 2 presents the methods used in this paper. Section 3 presents the study area and data while Section 4 sets out the obtained results. The concluding remarks are presented in Section 5.

2. Methods

The methods adopted in the current work are divided into three main topics: data mining, ordinary kriging and statement of the optimization problem.

2.1 Data mining using geo-regression

The analysis adopts (a) the average specific annual module as a primary study variable; (b) the coordinates of basin centroids as a basis to estimate the spatial variability structure, similarly to Merz and Blöschl (2005) (c) the drainage area as the proxy of a specific runoff.

The method requires defining a number M of evaluation basins and a number C of candidate basins as well a set of initial guesses. Because the M evaluation basins, the C candidate basins and the initial controlled basins are of various sizes, it is necessary to reduce the scale effect of drainage area. Assuming that the average specific annual module for a basin of size A_N is Q_N and assuming the scaling relationship $Q_N/Q=(A_N/A)^{\beta}$, the average specific annual module Q is replaced by the standardized specific module Q_N following Merz and Blöschl (2005) who adopted A_N= 100 km². It comes:

$$Q_N = \left(AA_N^{-1}\right)^{\beta} Q \tag{1a}$$

where Q_N is the average specific annual module for a hypothetical 100 square km basin, A (km^2) the gauged drainage area and Q is the observed average specific annual module. The scaling exponent β is found by a regression analysis between log(Q_N) and log(A). In Skøien et al. (2006), fitting resulted in β=-0.33 for mean annual discharge for Austria. To estimate β, several values are tested. Logarithmically transformed specific discharges are plotted as a function of the logarithm of drainage area for each tested β value to help verifying the model adequacy visually. Besides, the regression coefficient of determination R^2 is assumed as a quality criterion. Other criteria such as Root Mean Square Errors (RMSE) (Fair, 1986) or Akaike Information Criteria (AIC) (Bozdogan, 2000) can be assumed for model evaluation. However, R^2 is selected as an alternative to RMSE for it is dimensionless. The use of AIC is not needed because the number of parameters to be estimated is fixed regardless of the model (It is β which is to be estimated).

Moreover, the alternative of linking mean basin runoff to mean basin rainfall instead to drainage area is tested. A model similar to Eq. (1a) is proposed.

$$Q_N = \left(PP_{ref}^{-1}\right)^{\beta} Q \tag{1b}$$

where Q_N is the average specific annual module for a hypothetical reference rainfall of 100 mm, P (mm) the mean annual basin rainfall, P_{ref} (mm) is a reference rainfall and Q is the observed average specific annual module.

The best model (Eq. (1a) or Eq. (1b)) is finally selected on the basis of the performance measure R^2.

2.2 Interpolation using ordinary kriging

Kriging is a spatial interpolation method which takes into account the spatial variability of the data. This interpolation method is an unbiased estimator where the kriging (interpolation) error variance is minimized (Matheron, 1970). The basic idea of Euclidian kriging methods (such as ordinary kriging) is to estimate the value of a regionalized variable Z by a linear combination of the neighboring observations. Here, the neighboring observations are the basins which are "close" with respect to their centroid location when considering the prediction error Z of the fitted regression as a kriging variable.

The semivariogram is the structure function used here to model variability associated with the regionalized variable, Z. It measures the spatial variability of squared differences between pairs of variables, which allows building the experimental semivariogram, $\gamma(h)$, given by:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(x_i + h) - Z(x_i) \right]^2 \tag{2}$$

where x_i and x_i+h are two sampling locations separated by a distance h, $N(h)$ represents the number of sample points using h, $Z(x_i)$ and $Z(x_i+h)$ represent values of the variable Z measured at both locations. In this study, the basin centroids are adopted to locate basin sampling locations and determine the lags h as reported in Merz and Blöschl (2005). They are estimated by using GIS as reported in the Data section. The variable Z is related to the average specific annual discharge.

A semivariogram model is fitted to the experimental semivariogram. The fitted semivariogram is characterized by three main parameters: range, sill, and nugget. ‘Range’ represents the distance limit beyond which the data are no longer correlated. ‘Sill’ represents the variable variance. The ‘Nugget’ effect is a random component of the field Z and it represents either measurement errors or the variation of the studied variable at a small scale (Cressie, 1993).

Two semivariogram models are used here as alternatives: exponential and spherical. The exponential model for the semivariogram is given by Eq. (3):

$$\gamma(h) = \omega * \left[1 - \exp\left(-h/a\right)\right] + \omega_0 \tag{3}$$

where ω is the structural variance, ω_0 is the nugget variance and a is the range parameter. In the case of the exponential model, the range is defined as the distance at which the semivariogram is of 95% of the sill. So, it is equal to $3a$ according to Eq. (3). The sill is equal to ($\omega + \omega_0$) (Bardossy, 1997).

The spherical model for the semivariogram is given by Eq. (4):

$$\gamma(h) = \omega * \left[1.5 * \left(h/a\right) - 0.5 * \left(h/a\right)^3\right] + \omega_0 \tag{4}$$

Ordinary kriging is furthermore adopted. Thus, the estimated value $Z^*(x_0)$ at a location x_0 is a weighted linear combination of observations x_i at neighboring gauged basins i=1,N_{nb} where N_{nb} is the number of observations within the exploring neighbourhood (Matheron, 1970):

$$Z^*(x_0) = \sum_{i=1}^{N_{nb}} \lambda_i Z(x_i) \tag{5}$$

where $Z^*(x_0)$ is the estimated value of Z at the ungauged location x_0, λ_i is the weight given to the observation at the location x_i.

The variable Z in Eq. (2) and Eq. (5) is stated as the error (or regression residual) between the logarithms of the observed Q_N and the logarithms of the estimated Q_N by the regression model of Eq. (1a) or Eq. (1b).

The kriging weights λ_i are estimated as the solution of the ordinary kriging system (Eq. (6)):

$$\begin{cases} \sum_{i=1}^{N_{nb}} \lambda_j \gamma(x_j - x_i) + \mu' \\ \quad = \gamma(x_j - x_0) \quad j = 1,\ldots,N_{nb} \\ \sum_{i=1}^{N_{nb}} \lambda_i = 1 \end{cases} \tag{6}$$

where μ' is a Lagrange parameter accounting for the constraints on the weights (their sum is equal to unity). The x_j and x_i are the coordinates of the basin centroids, and $\gamma(x_j-x_i)$ is the estimated semivariogram for the lag between basin centroids x_j and x_i, using the theoretical semivariogram model. Thus, the weights λ_i and the Lagrange parameter μ depend entirely on the semivariogram model.

The kriging variance $\sigma_0{}^2$ helps to define and quantify the optimization objective function (Cressie, 1993; Barca et al., 2008). It is expressed for any ungauged location x_0 using the semivariogram model by:

$$\sigma_0{}^2 = \gamma(0) - \sum_{i=1}^{N_{nb}} \lambda_i \gamma(x_i - x_0) - \mu' \tag{7}$$

As stated earlier, the sample semivariogram is fitted to an exponential model and to a spherical model. The model parameters are evaluated by manual calibration. In fact, the first guess for each parameter is graphically adjusted. The acceptability of the fitted semivariogram model is then tested through the leave-one-out cross-validation scheme. This method removes a single data point, just one at a given time, and it estimates the result at the now missing location. The quality of the prediction is then evaluated. The parameter values are thus modified in order to obtain the best cross-validation results. The leave-one-out cross-validation is considered as one of the most commonly used methods to make an informed decision as to which model will provide the best predictions **(**Lin and Chen, 2004).

The standardized error and the coefficient of determination are adopted as criteria to evaluate the cross-validation results. The standardized error is equivalent to the value of the residuals between the observed values and the kriged Z^* values, divided by the standard deviation of kriging errors (Glatzer and Muller, 2004). Standardized residuals which are more than 2 and less than -2 are usually considered too large and, consequently, the parameters of the model semivariogram are modified in order to insure an acceptable range for the standardized residuals. The coefficient of determination (R^2) is also used for cross-validation (Laaha et al., 2014).

After performing the selection and validation of the fitted semivariogram model, the dependency ratio, which represents the percentage of the nugget effect (ω_0) in relation to the sill ($\omega + \omega_0$), is determined according to Cambardella et al. (1994). This ratio is used to interpret the strength of the dependency reported by the semivariogram structure. The higher the ratio is, the higher is the independency of the field observations. The values of dependency ratio are grouped and interpreted as follows: high dependency (< 25%), moderate dependency (25% - 75%), and low dependency (> 75%).

2.3 Statement of the optimization problem

2.3.1 Network design problem: Minimizing the average kriging variance

The problem statement is to extend an existing hydrometric network in order to evaluate the average specific annual module more accurately in the study basin. Thus, the optimization problem consists in minimizing an objective function defined here as the average kriging variance of error over a fixed evaluation grid, composed by i=1,M evaluation basins. This criterion, based on the geostatistical estimation error, is mentioned by Cressie (1993) among the criteria to adopt in network design problems:

$$OF = \sum_{i=1}^{M} \sigma_i^{\,2} \Big/ M \tag{8}$$

This objective function depends entirely on the semivariogram model and on the *M* selected grid points. The collection of the *M* centroids constitutes what is called the “grid nodes”. For kriging implementation, these evaluation basins are required to be different from the controlled basins.

Thus, when a “grid” of *M* basins is adopted to compute the variance of kriging error and quantify the objective function *OF*, the minimization problem is solved by using a simulated annealing algorithm (Kirkpatrick et al., 1983). Indeed, the simplicity of the algorithm and the variety of optimization problems to which the algorithm is used are among the main advantages of simulated annealing (Fleischer, 1995). This algorithm is applied in Cunha (1999) for solving aquifers' management problems. It was also applied by Chebbi et al. (2011) in order to optimize the selection of rainfall stations in the issue of increasing the size of an existing rainfall network.

2.3.2 Definition of candidate solutions and simulation scenarios

The optimal locations are chosen from the *C* candidate stations which are represented by the centroids of their drainage area. The candidate stations are selected in such a way that they cover the whole study region. Besides, they are selected in such a way that they do include outlets representing upstream basins, and small to moderate size basins. Moss and Tasker (1991) recommended that the number of candidate stations should be at least three times the number of the desired optimal stations. In this work, due to the high cost of the hydrometric equipments and to the financial constraints, we seek to implement only one to five new stations. Thus, the new locations investigated by using the SA optimization scheme for five scenarios respectively involve: (1) a network consisting of 13 hydrometric gauges, (2) a network consisting of 14 hydrometric gauges, (3) a network consisting of 15 hydrometric gauges, (4) a network consisting of 16 hydrometric gauges, and (5) a network consisting of 17 hydrometric gauges, including all N=12 existing stations. The same 15 candidate locations are investigated for the five scenarios in order to allow for an inter-comparison scenario, whatever the final size of the optimized network.

3. Study area and data

The study area is the North of Tunisia, including the Medjerda basin (BV5), the Northern Coast Basin (BV3) and the Cap Bon – Méliane Basin (BV4). However, the optimization has been performed for the Medjerda Basin (BV5) which covers an area of 21,000 km^2 in Tunisia. Figure 1 shows the hydrometric network of the study area composed of 19 controlled basins. Their names and drainage areas are reported in Table 1. Twelve out of the 19 controlled basins are located in Medjerda Basin (BV5) and the remaining are in its neighboring basins. Six stations are located in the North Coast Basin (BV3) and one single station is part of the Cap Bon – Méliane Basin (BV4). Neighboring basins belonging to BV3 and BV4 are used both for developing the spatial variability assessment during the sample semivariogram estimation and for kriging in the cross validation step. In addition to the twelve stations studied in Medjerda Basin (BV5), two other stations are located in the Tessa sub basin but are not included in the sample. They are Pont Route Souani on Oued Souani, a tributary of Tessa and Sidi Mediane on Oued Tessa. The reason is that the observed average annual modules of these two stations have singularities. In addition, Oued Souani is already controlled by a dam achieved since 2005. This is why these two stations are not taken into consideration in the initial hydrometric network of Medjerda.

In this work, all basin boundaries are derived from a digital elevation model available within a 30-meter resolution (ASTER, 2012). Furthermore, the coordinates of the basin centroids are derived from the resulting basin boundaries using ArcGIS. Sizes of the 12 gauged basins of the Medjerda basin range from 60 to 20811 km^2.

A brief description of the Medjerda tributaries is required to understand the motivation that lies behind the selection of the *M* evaluation basins (grid nodes) and the *C* candidate basins.

In the right bank of the Medjerda river, (viewed from upstream direction), the main direct tributaries are Oued (river) Mellegue which is partly situated in Algeria, Oued Tessa, Oued Siliana and Oued Lahmar. On the left bank, the main tributaries are Oued Rarai, Oued Bouheurtma, Oued Kasseb, Oued Beja and Oued Zerga. Oued Mkhachbia is a very small basin neighboring Oued Beja. Tributaries of the right bank are much longer and steeper than those of the left bank and they are much subjected to water erosion. On the other hand, some of the right bank tributaries, such as Rmil, a tributary of the Siliana river and Rmel, a tributary of the Mellegue river, are responsible for intense floods (Rodier et al., 1981; Ghorbel, 1997; Zahar et al., 2008). Thus, the selection of M and C basins requires considering the basin location: left bank or right bank.

Table 1 displays the observed average annual module which is reported by using the National hydrological service (DGRE) annual reports. Figure 1 and Table 1 show that there is a lack in the observation of upstream sub basins of the Medjerda River. Indeed, historically speaking, this network is aimed to design the existing large dams. Besides, it is intended for flood forecasting purposes. This might explain why small basins are left aside in the current network conception. Thus, network size augmentation may help to correct this kind of bias in the drainage area coverage. So, the selection of M and C basins needs to include basins of small and moderate sizes.

Because we deal with one to five new sites, it is assumed that M=20 is sufficient to compute the grid average kriging error with confidence. The sampling of evaluation grid basins is conceived in such a way as to cover the study domain (in both left and right banks) and to include small, moderate and large drainage areas. Figure 2 shows the “grid” node locations of the M=20 basins selected for the evaluation of the objective function. On the other hand, for the purpose of successively selecting one to five new basins to be controlled, 15 candidate locations are selected (Figure 3). Similarly, candidate stations are chosen on either the right or the left bank. For example, for Tessa basin on the right bank, two candidate sites (C2 and C14), are prospected respectively upstream and downstream of an important river recharge area (Figure 3). For Zerga basin on the left bank, two candidate sites are also proposed (C9 and C10) as their tributaries meet at a confluence (Figure 3). Table 2 reports the basin size and the tributary of the 15 candidate stations as well as the description of the reason of their selection. As needed to improve the network cover for small and moderate basins sizes, the candidate drainage areas vary from 107 to 755 km².

To adjust the geo-regression parameter β, when using the drainage area as attribute (Eq. (1a)), a network of 39 well-documented gauged basins belonging to the National hydrometric network of Tunisia is considered. Their sizes vary from 3 to 20811 km². Their average annual modules vary between 0.05 and 27.5 m^3/s. The plot of the logarithms of the observed average specific annual module versus the logarithms of the drainage area is reported in Figure 4 where the 19 basins of Northern Tunisia are made distinguishable from the whole sample of 39 basins. For the other alternative of linking mean basin runoff to the mean basin rainfall (Eq. (1b)), only a subgroup of 21gauged basins, among the existing 39 ones, is used to adjust the geo-regression parameter β. In fact, mean annual rainfall data are available only for these 21 stations.

4. Results

4.1 Scaling and regression results

The scaled specific discharge Q_N sample (Eq. (1a)) is estimated for various hypothetical β values using the 39 stations. The best estimator of the exponent β is achieved for $\beta=-1.5$ according to R^2. The Ln-Ln linear regression relation is reported in Figure 5a. It results in $R^2>0.8$, reflecting a good performance. The alternative of linking mean basin runoff to mean basin rainfall rather than to drainage area (Eq. (1b)) results in $\beta=0.1$ with P_{ref} =100 mm as the most appropriate estimation. The Ln-Ln linear regression relation is reported in Figure 5b.

The coefficient of determination R^2 is equal to 0.79, which is less satisfactory than the R^2 obtained using drainage area (0.89). Thus, we further assume the drainage area as the sole attribute.

The scatter plot of the residuals against the explanatory variables (logarithm of drainage areas) is now examined. Figure 5c shows no decrease or increase of residuals with the increase in the logarithm of drainage areas, thus revealing no heteroscedasticity of the variable Z (errors). The resulting values are shown in Table 3.

4.2 Spatial variability results

The residuals of regression estimation of the average specific annual modules Q_N in the 19 gauged basins are assumed as a variability pattern Z to be analyzed and to be used to quantify the sample semivariogram. The latter is reported in Figure 6 as well as the size of the samples which are used to derive it.

The fitted exponential model is without any nugget effect, displaying a range parameter of 30 km and a sill parameter of 1.4 $(m^3/s/km^2)^2$. The fitted spherical model is without any nugget effect, with a range of 50 km and a sill parameter of 1.2 $(m^3/s/km^2)^2$. For these two models, the dependency ratio is equal to 0, which translates a strong spatial dependency in the data. This is well-understood since residual errors originate from regression using mean squares errors with unbiased mean error.

The exponential semivariogram model yields satisfactory cross validation results since the standardized errors are all varying in the acceptable interval range [-2, 2] (see Table 3). Besides, the determination coefficient R^2 is equal to 0.72 which is nearly the value obtained in Laaha et al. (2014) for top-kriging (R^2=0.75). For the spherical model, the cross-validation results are less convincing than those obtained with the exponential model. For instance, for the Mkhachbia station (O4), the standardized error is less than -2 (see Table 3). Besides, the determination coefficient R^2 is equal to 0.45, namely much lower than that of the exponential model.

Thus, the exponential model is adopted as a spatial variability structure since it gives the best results in cross-validation.

4.3 Augmented hydrometric networks results

As presented in the methodology, to achieve the optimization objective, the spatial average kriging variance of the interpolation error Z is minimized over the candidate networks using simulated annealing. As expected, we notice that, as the network size goes up, the estimation average variance goes (Table 4), thus reflecting the increase in spatial interpolation accuracy with the increase in the number of network hydrometric gauges. Moreover, seemingly, the number of additional stations may still go up since the curve relating spatial average kriging variance to the number of additional stations has not reached a sill. In this work, it is assumed a maximum of 5 new stations only because of financial constraints. It seems that this size can be increased as the optimal size of the network has not been reached yet.

As an example, Figure 7 shows the spatial distribution of an optimized network for Scenario 5. The five new stations are spared between the left and right banks. One basin upstream at the Algerian frontier is selected (C15). Various basin sizes are covered by the selected locations ranging from 209 to 594 km^2 while the range for the candidate locations is from 107 to 755 km^2. Indeed, the selected stations are distributed adequately around the Medjerda River. It seems that the algorithm operated a synthesis in both upstream and downstream directions, as well as between the left bank and the right bank.

The resulting optimal stations obtained from the five scenarios are listed in Table 5. From a scenario to another, it is worth noticing that there is a perfect conformity with regards to the new sites when progressing from 1 to 5 stations. This means that, for a given scenario, the locations of the new sites include the optimal stations which have already been chosen in the previous scenario. This indicates the robustness of the location identification and its practical importance.

4.4 Interpretation

The presentation of the results is as follows: in Scenario 1, the selected station is the candidate C14 located upstream on the Tessa's tributary (right bank) with a 509 square kilometer basin. In Scenario 2, in addition to station C14, the optimization indicates that a station (C4) should be implemented downstream in Lahmar tributary (right bank) for a 594 square kilometer basin. In Scenario 3, the previously selected stations (C14 and C4) are also reselected and the third location is recommended on Beja tributary, on the left bank (candidate C5) with 209 square kilometer basin. In Scenario 4, the selection of the previous stations (C14, C4 and C5) is confirmed and the fourth station is recommended on the Rmil tributary of Siliana River in the right bank (C7) for a 277 square kilometer basin. Finally, in Scenario 5, the four previous recommended stations (C14, C4, C5 and C7) are maintained with an additional basin C15 on the left bank, located on a Medjerda's tributary at the Algerian frontier with a 245 square kilometer drainage area. This last new station is proposed for the upper stream near the river course, far from the first four selections.

What are the implications of the findings with respect to the average inter-station distance, average drainage area as well as minimum and maximum inter-stations distances?

Table 5 reports the average inter-station distances as well as the average drainage area for each scenario, together with the minimum and maximum inter-stations distances. The lowest minimum inter-stations distance (about 11 km) is given by the initial network of 12 stations. As no candidate is proposed with a smaller inter-distance, the minimum remains unchanged.

For Scenario 1 (adding one single station), maximum inter stations distance remains that given by the initial network of 12 stations (about 168 km). In fact, this maximum value is the distance between Mkhachbia and Mellegue K13 basins (from respectively the East side and the West side of the Medjerda basin). Figure 8 shows the progression in mean centroids inter-distances. The algorithm decreases the basin inter-distances when selecting one new location (Scenario 1). This insures a better spatial coverage. The addition of two best locations to the initial network is achieved in order to extend the network, which is reflected by the augmentation of the maximum basins inter-distance (Figure 8). The increase of a maximum inter-distance is achieved together with an increase in mean inter-distance in Scenario 2.

From Scenario (3) to Scenario (5) the average inter-station distance is increased and then decreased (Table 5), while conversely, the mean drainage area is regularly decreased from Scenario (1) to Scenario (5) (Table 5). This shows that the optimized networks keep candidates that shift the average drainage area of the optimized network towards higher specific discharges ranges.

5. Conclusions

An approach based on geo-regression combined to ordinary kriging of log specific runoff versus log drainage area residuals is adopted to extend a hydrometric network in order to evaluate an important hydrological descriptor, the average specific annual module, more accurately. To achieve the optimization objective, the spatial average kriging variance of the kriging interpolation error is considered. The kriged variable is the error of estimation of the normalized (scaled) average specific discharge by regression using drainage area.

The minimization of the objective function represented by the mean areal variance of kriging error is achieved by using simulated annealing.

The Northern region of Tunisia, which has a sub-humid to semi-arid climate, is used in order to develop the methodology. The approach is based on the evaluation of five scenarios for augmenting the size of an initial network of 12 stations. The analysis of the optimized locations from a scenario of one single additional station to five additional stations shows a perfect agreement in relation to the new sites' location. Actually, the locations of the new sites include the optimal stations already chosen in the previous scenarios. The new locations insure a better spatial coverage of the study area as seen from the increase of the average and the maximum inter-station distances after optimization. The results also show that the optimized networks introduce basins that insure the shifting of the mean drainage area towards higher specific discharges ranges. There is no limitation to apply this kind of study elsewhere provided that a significant link exists between the drainage area and the specific mean runoff, and also, provided that a scaling formula may be fitted. In the absence of a significant link between the drainage area and the specific discharge, other proxy variables should be selected. If the scaling formula could not be fitted for the whole study area, a regionalization of the scaling formula is recommended. The type of semivariogram model (exponential) selection is not considered as a limitation. In fact, the only limitation is that the optimization should be performed in accordance with the range of the semivariogram (the location of the new sites must respect the de-correlation distance of the fitted semivariogram). The perspectives in research topics aim to develop a multi-objective optimization problem so that it can include the financial concerns and the optimal size of the network. Besides, the method of Particle swarm optimization (Taormina and Chau, 2015) is proposed as a perspective for the optimization algorithm.

REFERENCES

Acreman, M.C., Sinclair, C.D., 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. J. Hydrol. 84, 365–380.

Adamowski, K., Bocci, C., 2001. Geostatistical regional trend detection in river flow data, Hydrol. Processes. 15, 3331–3341.

ASTER Global Digital Elevation Map Announcement, 2012. https://asterweb.jpl.nasa.gov/gdem.asp.

Barca, E., Passarella, G., Uricchio, V., 2008. Optimal extension of the rain gauge monitoring network of the Apulian Regional Consortium for Crop Protection. Environ Monit Assess. 145, 375−386.

Bardossy, A., 1997. Introduction to geostatistics. Technical report, Institute for Hydraulic Engineering and Water Resources Management, University of Stuttgart.

Bozdogan, H., 2000. Akaike's information criterion and recent developments in informational complexity. Journal of Mathematical Psychology. 44, 62−91.

Bum, D.H., Goulter, I.C., 1991. An approach to the rationalization of streamflow data collection networks. Journal of Hydrology. 122, 71−91.

Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F., Konopka, A.E., 1994. Field-scale variability of soil properties in central Iowa soils. Soil Science Society of America Journal. 58, 1501−1511.

Caselton, W.F., Husain, T., 1980. Hydrologic network: Information transmission. J. of Water Resour. Plan. and Manag. Div. 106(2), 503−520.

Castellarin, A., 2007. Probabilistic envelope curves for design flood estimation in ungauged sites. WRR 43(4), W04406 1−12.

Chebbi, A., Kebaili Bargaoui, Z., Da Conceiçao Cunha, M., 2011. Optimal extension of rain gauge monitoring network for rainfall intensity and erosivity index interpolation. Journal of Hydrologic Engineering. 16(8), 665−676.

Chebbi, A., Kebaili Bargaoui, Z., Da Conceiçao Cunha, M., 2013. Development of a method of robust rain gauge network optimization based on intensity-duration-frequency results. Hydrology and Earth System Sciences. 17, 4259−4268.

Cressie, N.A.C., 1993. Statistics for Spatial Data, (Revised Edition). Wiley: New York.

Cunha, M.C., 1999. On solving aquifer management problems with simulated annealing algorithms. Water Resources Management. 13, 153−169.

Daigle, A., St-Hilaire, A., Beveridge, D., Caissie, D., Benyahya, L., 2011. Multivariate analysis of the low-flow regimes in eastern Canadian rivers. Hydrological Sciences Journal. 56(1), 51−67.

Davar, Z.K., Brimley, W.A., 1990. Hydrometric network evaluation: Audit approach, J. Water Resour. Plann. Manage. 116, 134−146.

Daviau, J-L., Adamowski, K., Patry, G.G., 2000. Regional flood frequency analysis using GIS, L-moment and geostatistical methods. Hydrol. Processes. 14, 2731−2753.

Eaton, B., Church, M., Ham, D., 2002. Scaling and regionalization of flood flows in British Columbia, Canada. Hydrol. Processes. 16, 3245−3263.

Fair, R.C., 1986. Chapter 33 Evaluating the predictive accuracy of models. Handbook of Econometrics. 3, 1979−1995.

Fleischer, M.A., 1995. Simulated annealing: past, present, and future. In: C. Alexopoulos, K. Kang, W.R. Lilegdon and D. Goldsman (eds.), Proceedings of the 1995 Winter Simulation Conference, IEEE Press, 155−161.

Ghorbel, A., 1997. Dynamique fluviale de la Medjerda : Identification des points de débordements depuis le barrage de SIDI SALEM jusqu'à la mer. DGRE, 11 p.

Glatzer, E., Muller, W.G., 2004. Residual diagnostics for variogram fitting. Computers and Geosciences. 30, 859−866.

Hundecha, Y., Bardossy, A., 2004. Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. Journal of Hydrology. 292, 281−295.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by Simulated Annealing. Science. 220(4598), 671−680.

Kjeldsen, T.R., Jones, D.A., 2007. Estimation of an index flood in the UK, Hydrol. Sci. J. 52, 86−98.

Kron, W., Willems, W., 2002. Flood risk zoning and loss accumulation analysis for Germany, International Conference on Flood Estimation, 6−8 March 2002, Berne, Switzerland, pp. 549−558.

Laaha, G., Skøien, J.O., Blöschl, G., 2014. Spatial prediction on river networks: comparison of top-kriging with regional regression. Hydrol. Process. 28, 315−324.

Lin, G.F., Chen, L.H., 2004. A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. Journal of Hydrology. 288, 288−298.

Matheron, G., 1970. Cours de géostatistique, chapitre VI. Notes de cours de DEA.

Merz, R., Blöschl, G., 2005. Flood frequency regionalization—Spatial proximity vs. catchment attributes. J. Hydrol. 302, 283−306.

Mishra, A.K., Coulibaly, P., 2009. Developments in hydrometric network design: A review. Reviews of Geophysics. 47(2), RG2001 / 2009, 1−24.

Moss, M.E., Tasker, G.D., 1991. An intercomparison of hydrological network-design technologies. Hydrological Sciences Journal. 36, 209−221.

Rodier, J.A., Colombani, J., Claude, J., Kallel, R., 1981. Monographies hydrologiques, le bassin de la Medjerdah, ORSTOM, Direction des resources en eaux et en sols, Division des ressources en eau, Service hydrologique de Tunisie, 472p.

Rouhani, S., 1985. Variance reduction analysis. Water Resources Research. 21(6), 837−846.

Sadegh, M., Vrugt, J.A., Gupta, H.V., Xu, C., 2016. The soil water characteristic as new class of closed-form parametric expressions for the flow duration curve. Journal of Hydrology. 535, 438−456.

Skøien, J.O., Blöschl, G., Western, A.W., 2003. Characteristic space-time scales in hydrology. Water Resour. Res. 39(10), 1304.

Skøien, J.O., Merz, R., Blöschl, G., 2006. Top-kriging – geostatistics on stream networks. Hydrol. Earth Syst. Sci. 10, 277−287.

Taormina, R., Chau, K.W., 2015. Data-driven input variable selection for rainfall-runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. Journal of Hydrology. 529 (3), 1617−1632.

Tarboton, D.G., Bras, R.L., Puente, C.E., 1987. Combined hydrologic sampling criteria for rainfall and streamflow, J. Hydrol. 95, 323−339.

Tasker, G.D., Stedinger, J.R., 1989. An Operational GLS Model for hydrologic regression. J. of Hydrol. 111(1-4), 361−375.

Vivekanandan, N., 2012. Evaluation of Stream Flow Network using Entropy Measures of Normal and Lognormal Distributions. Bonfring Journal of Industrial Engineering and Management Science. 2(3), 33−37.

Vivekanandan, N., 2014. Evaluation of Stream Flow Network Using Entropy. Current Advances in Civil Engineering (CACE). 2(3), 90−94.

Wilson, J.P., Gallant, J.C., 2000. Digital Terrain Analysis. Chap. 1 in Terrain Analysis: Principles and Applications.

Zahar, Y., Ghorbel, A., Albergel, J., 2008. Impacts of large dams on downstream flow conditions of rivers: Aggradation and reduction of the Medjerda channel capacity downstream of the Sidi Salem dam (Tunisia). Journal of Hydrology. 351, 318−330.

Figure 1

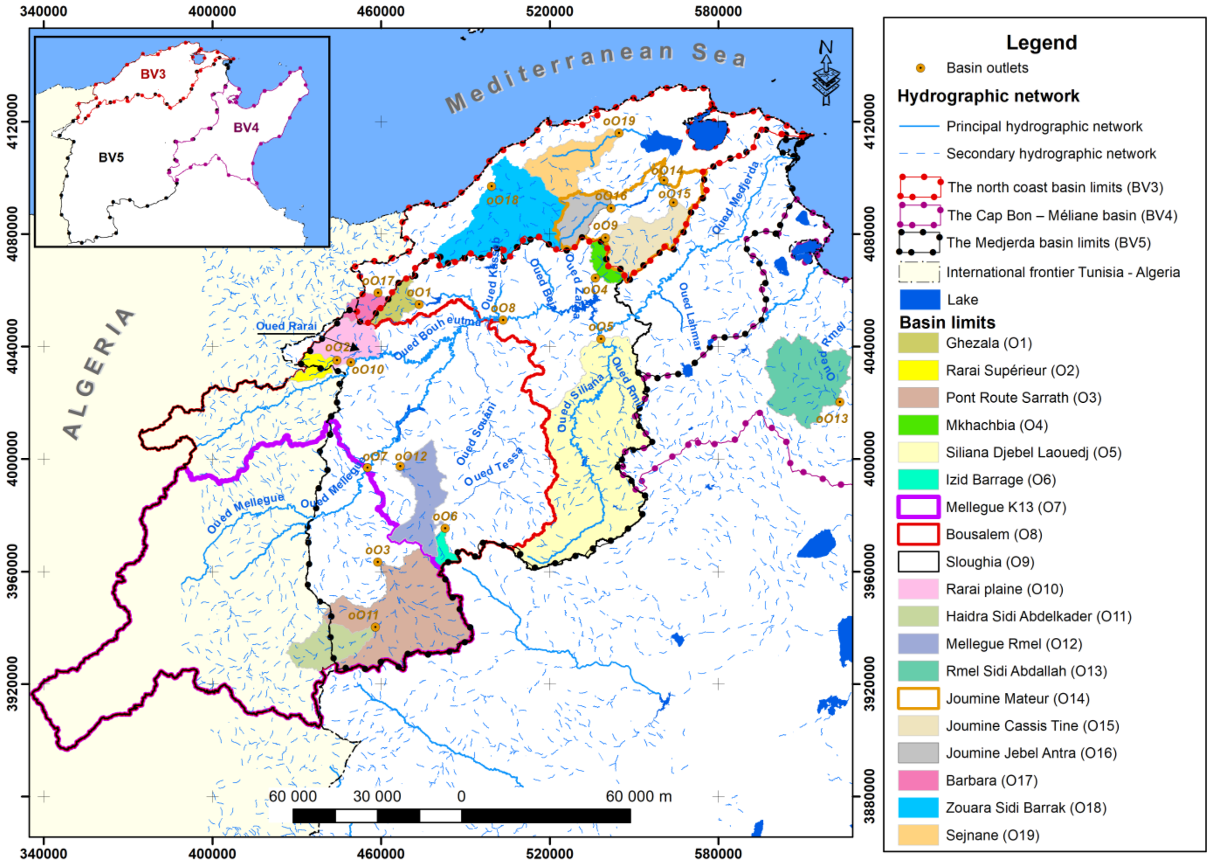

Figure 1. Gauged basin outlets for the 19 stations

Figure 2

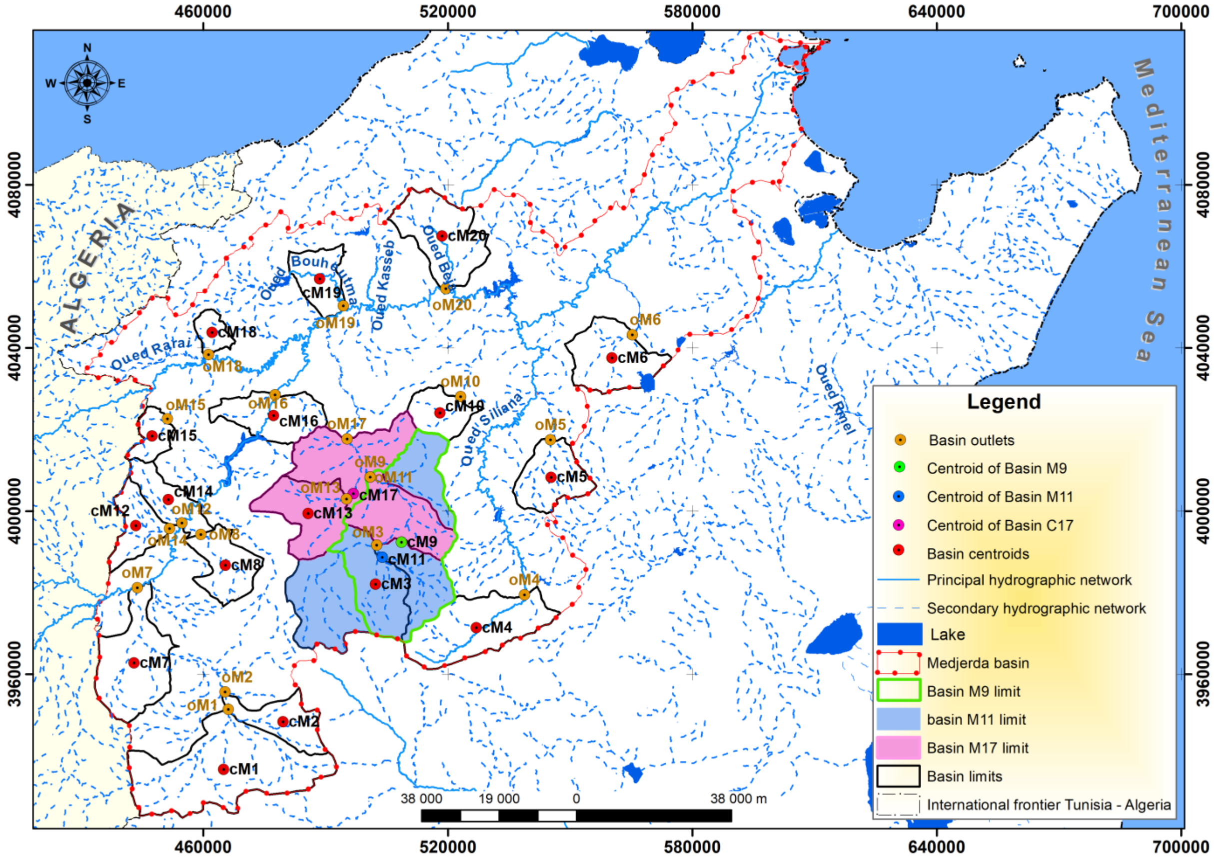

Figure 2. Basins centroids and oulets of “grid” nodes for M=20 selected basins

Figure 3

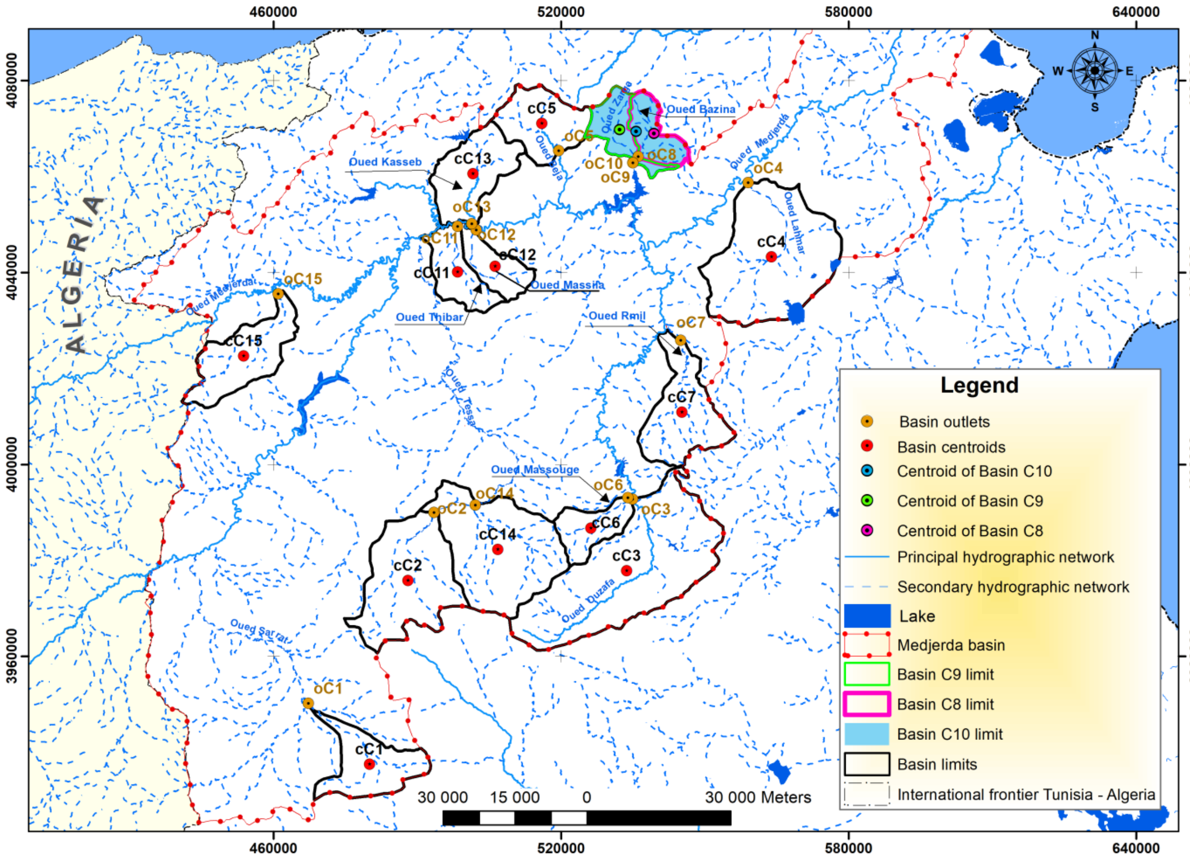

Figure 3. Centroids of the fifteen candidate locations for composing the optimal network

Figure 4

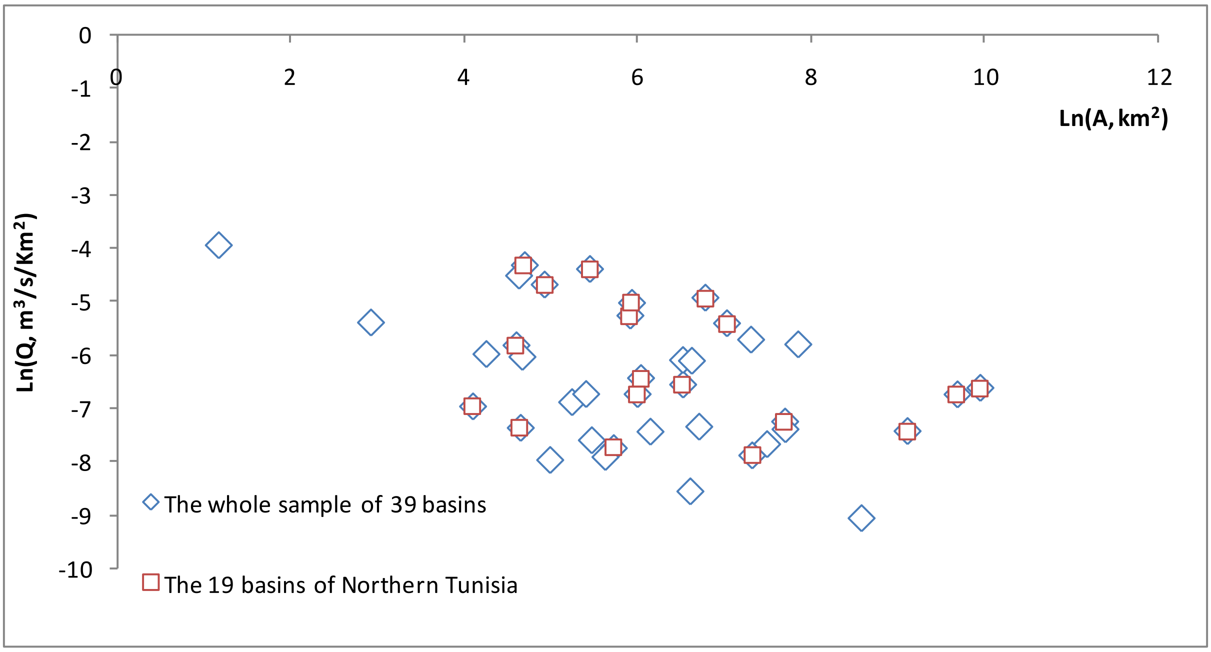

Figure 4. The logarithms of observed average specific annual module versus the logarithms of drainage area

Figure 5

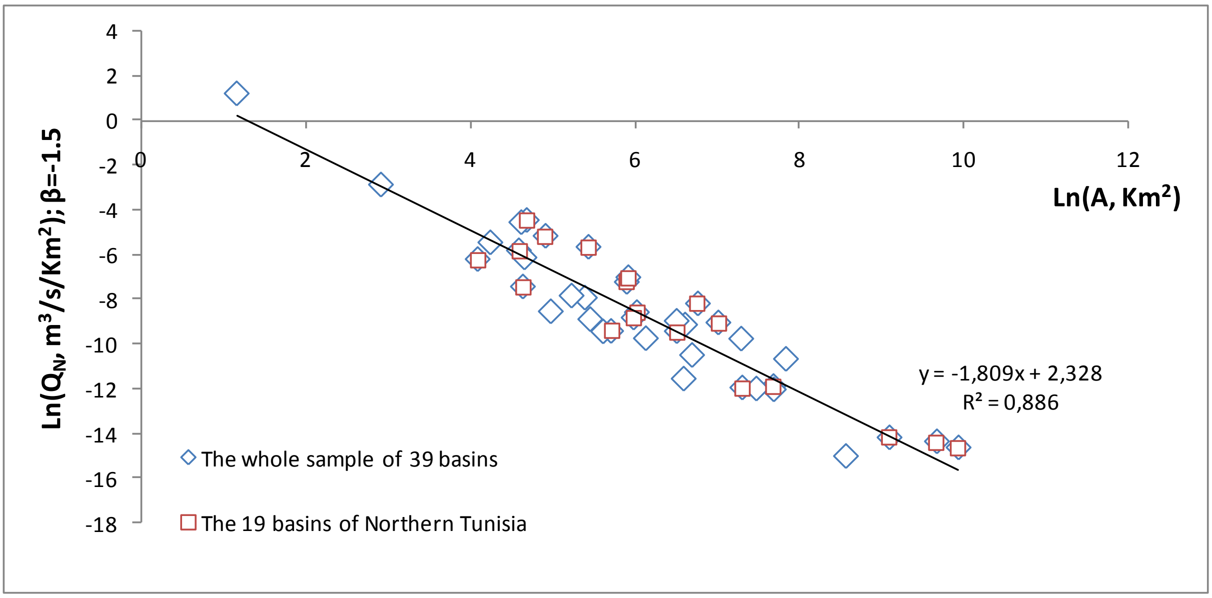

Figure 5a. Linear regression relation between the logarithm of scaled specific discharge Q_N and the logarithm of drainage area A.

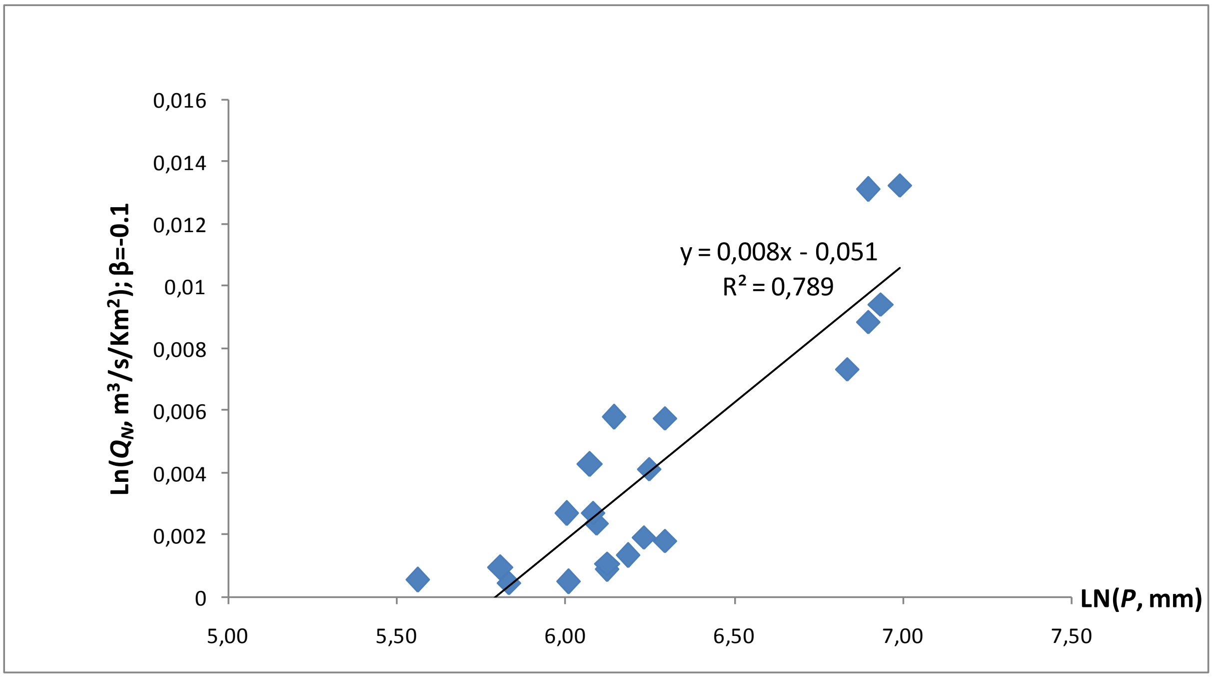

Figure 5b. Linear regression relation between the logarithm of scaled specific discharge Q_N and the logarithm of basin mean rainfall P (for 21 gauged basins with rainfall information)

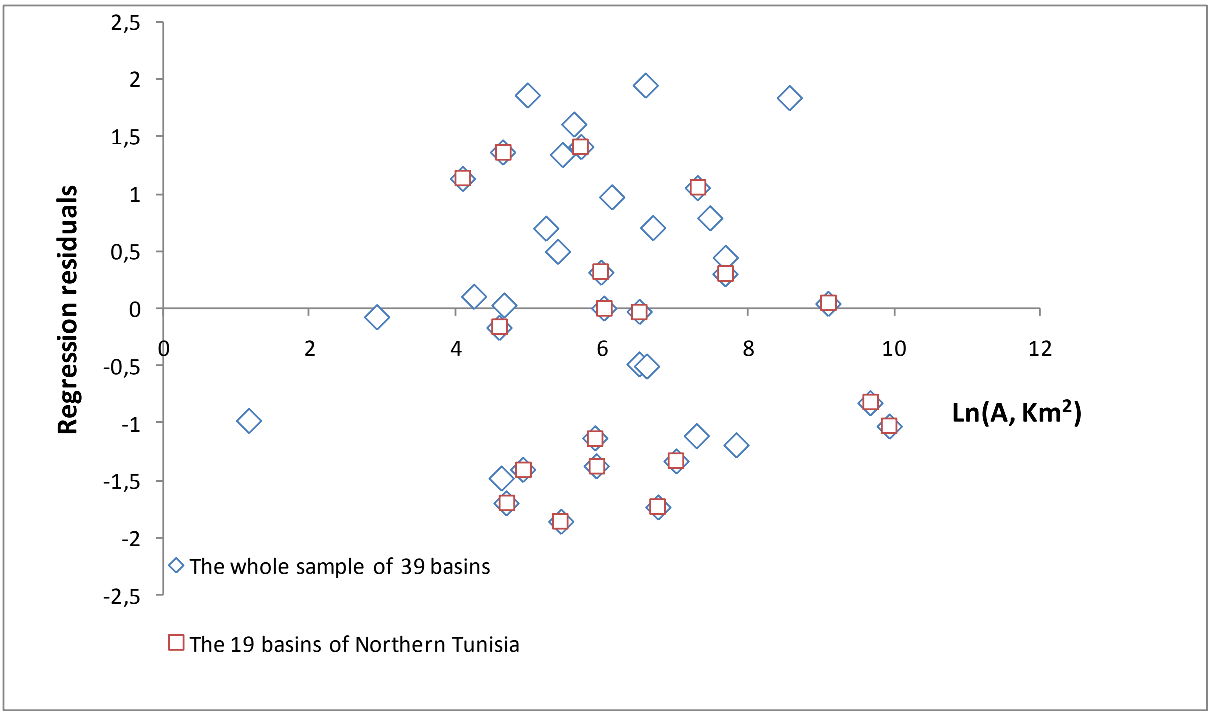

Figure 5c. Regression residuals versus logarithm of drainage areas

Figure 6

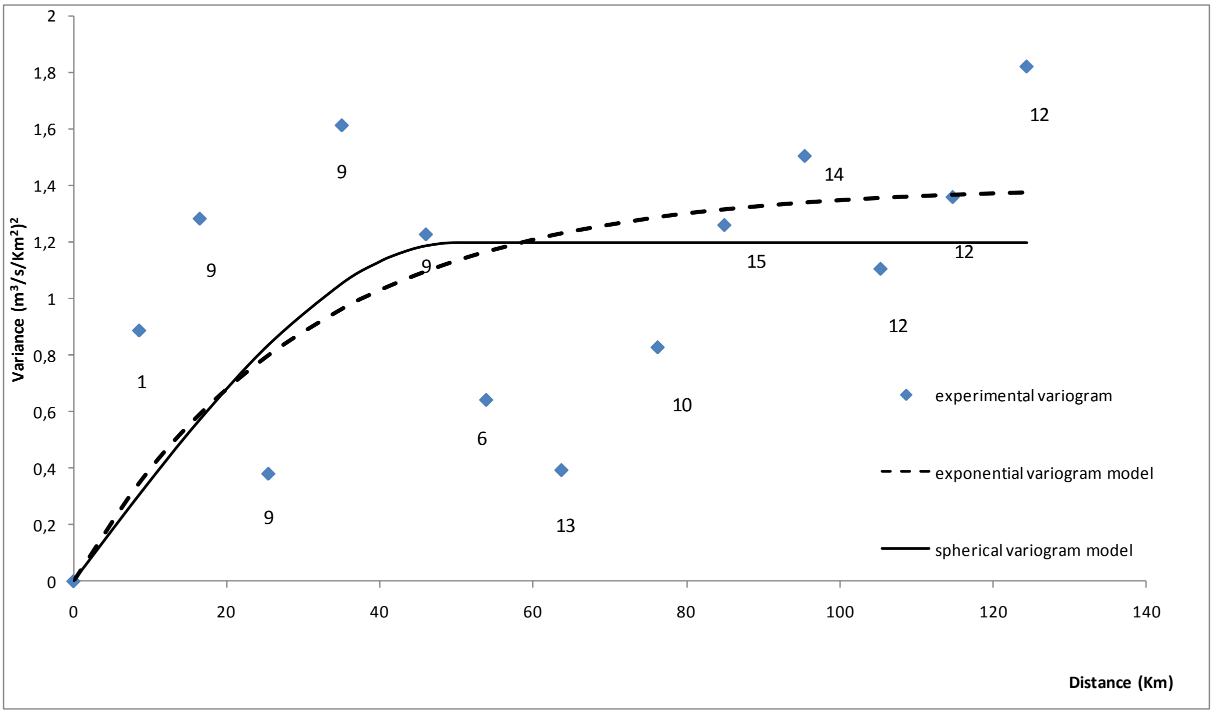

Figure 6. Calibration of the semivariogram of residuals of estimation of the scaled average specific discharge (with the corresponding sample size of pairs)

Figure 7

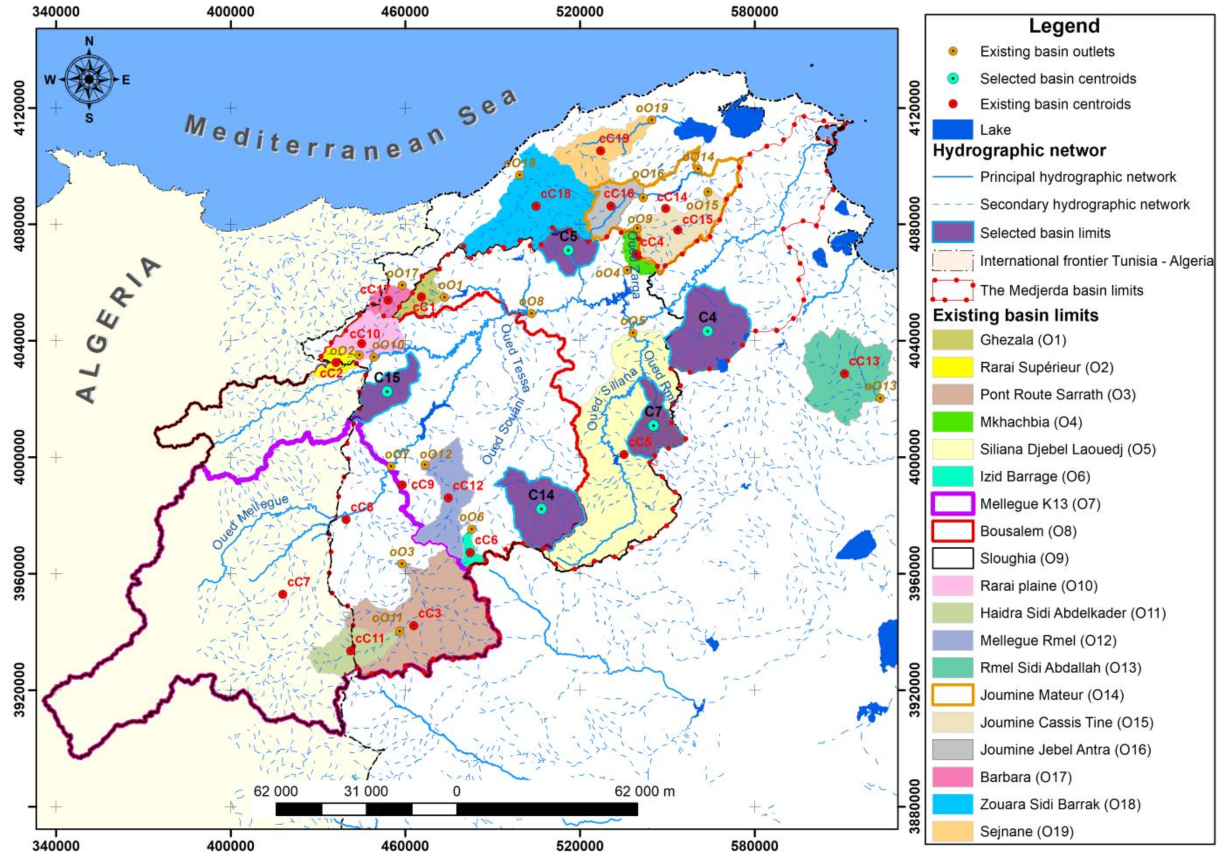

Figure 7. Spatial distribution of the optimized hydrometric network for Scenario 5 with 5 new sites.

Figure 8

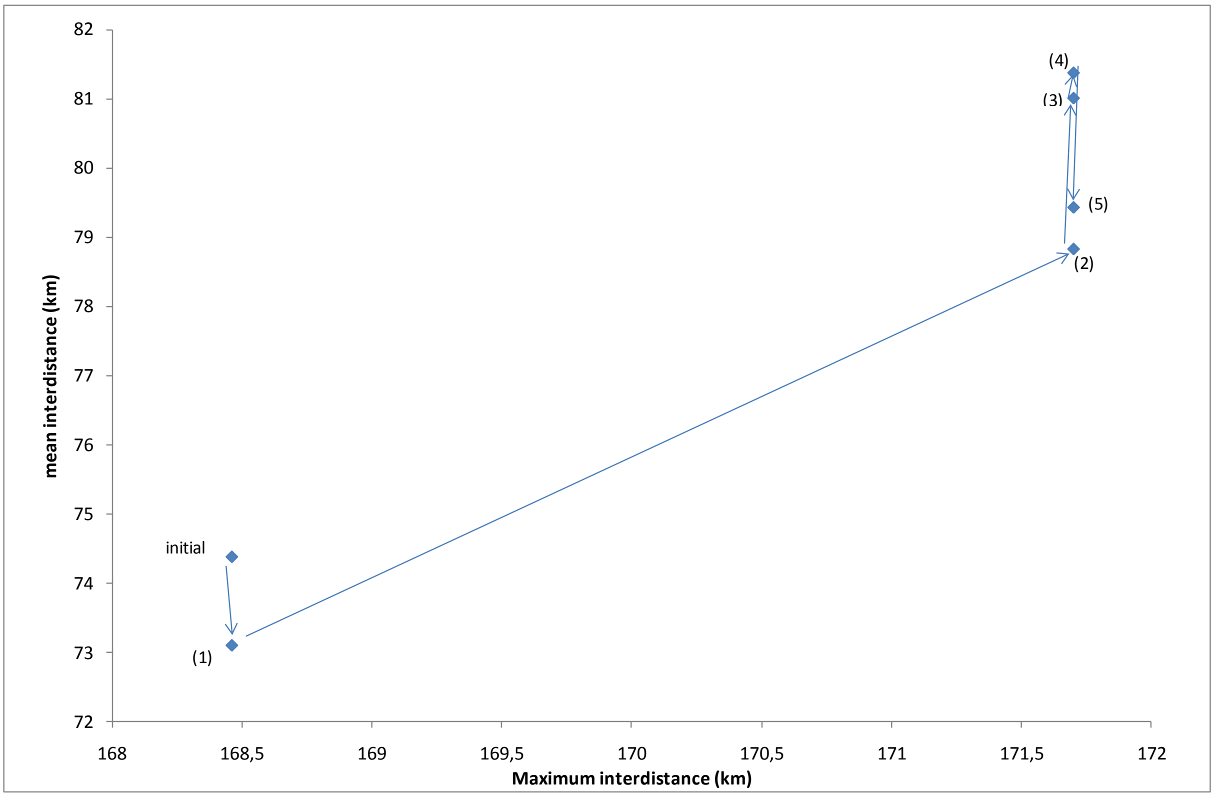

Figure 8. The progression in mean and maximum centroids interdistance according to the Scenario (1) to (5)

Table 1
Click here to download Table: Table 1.docx

Table 1. Presentation of the 19 hydrometric stations of the study area

Station name	Basin	Station Code in Fig. 1	Drainage area (km^2)	Average annual module (m^3/s)
Ghezala in Bouheurtma basin	BV5	O1	137	1.25
Rarai Supérieur in Rarai basin	BV5	O2	99	0.29
Pont Route Sarrath in Mellegue basin	BV5	O3	1500	0.56
Mkhachbia in Mkhachbia basin	BV5	O4	104	0.06
Siliana Djebel Laouedj in Siliana basin	BV5	O5	2191	1.54
Izid Barrage in Tessa basin	BV5	O6	60	0.06
Mellegue K13 in Mellegue basin	BV5	O7	8988	5.29
Bousalem along the Medjerda river	BV5	O8	15993	18.70
Sloughia along the Medjerda river	BV5	O9	20811	27.50
Rarai plaine in Rarai basin	BV5	O10	368	1.88
Haidra Sidi Abdelkader in Mellegue basin	BV5	O11	304	0.13
Mellegue Rmel in Mellegue basin	BV5	O12	400	0.47
Rmel Sidi Abdallah in Rmel basin, BV4	BV4	O13	676	0.95
Joumine Mateur in Joumine basin	BV3	O14	1121	4.96
Joumine Cassis Tine in Joumine basin	BV3	O15	416	0.66
Joumine Jebel Antra in Joumine basin	BV3	O16	231	2.82
Barbara in Barbara basin	BV3	O17	109	1.43
Zouara Sidi Barrak in Zouara basin	BV3	O18	874	6.24
Sejnane in Sejnane basin	BV3	O19	375	2.43

Table 2

Table 2. Presentation of the 15 candidate stations (area, tributary and location). RB: right bank, LB: left bank of Medjerda river

Station code	Drainage area (km^2)	Name of the tributary and description of location
C1	214	In upstream Sarrat river, in Mellegue basin, RB
C2	419	On Tessa river, RB
C3	755	On Ouzafa river, tributary of Siliana river, RB
C4	594	On Lahmar river, RB
C5	209	On Beja river, LB
C6	124	On Massouge river, tributary of Siliana river, RB
C7	277	On Rmil river, tributary of Siliana river, RB
C8	107	On Bazina river, tributary of Zarga river, LB
C9	112	On Zerga river, LB
C10	219	On Zerga river, LB
C11	147	On Massila river, LB
C12	116	On Thibar river, RB
C13	236	On Kasseb river, LB
C14	503	On Tessa river, RB
C15	245	Ezana direct tributary of the Medjerda basin, at the Algerian boundary, LB

Table 3
Click here to download Table: Table 3.docx

Table 3. Error for the selected variogram and cross-validation results: the standardized errors at the 19 stations of the study area

Station name	Station Code	Error Z for the selected variogram ($m^3/s/Km^2$)	For the exponential model	For the spherical model
Ghezala in Bouheurtma basin	O1	-1.40	0.68	0.10
Rarai Supérieur in Rarai basin	O2	-0.17	-0.47	-0.74
Pont Route Sarrath in Mellegue basin	O3	1.05	-0.48	-0.38
Mkhachbia in Mkhachbia basin	O4	1.36	-1.71	-2.44
Siliana Djebel Laouedj in Siliana basin	O5	0.30	-0.29	-0.60
Izid Barrage in Tessa basin	O6	1.13	-0.56	-0.84
Mellegue K13 in Mellegue basin	O7	0.040	0.25	0.12
Bousalem along the Medjerda river	O8	-0.82	-0.61	0.33
Sloughia along the Medjerda river	O9	-1.03	1.17	0.98
Rarai plaine in Rarai basin	O10	-1.13	0.70	0.59
Haidra Sidi Abdelkader in Mellegue basin	O11	1.41	-1.14	-1.12
Mellegue Rmel in Mellegue basin	O12	0.31	-0.79	-0.46
Rmel Sidi Abdallah in Rmel basin, BV4	O13	-0.03	0.20	-0.38
Joumine Mateur in Joumine basin	O14	-1.33	0.86	1.32
Joumine Cassis Tine in Joumine basin	O15	0.00	-0.46	-0.65
Joumine Jebel Antra in Joumine basin	O16	-1.86	0.80	1.41
Barbara in Barbara basin	O17	-1.70	-0.26	0.60
Zouara Sidi Barrak in Zouara basin	O18	-1.73	0.48	0.83
Sejnane in Sejnane basin	O19	-1.37	-0.72	-0.31

Table 4. Reduction of uncertainty by increase of network density

Increasing the existing network by	spatial average kriging variance $(m^3/s/km^2)^2$
Initial 12 stations network	0.92
1 new station	0.82
2 new stations	0.78
3 new stations	0.73
4 new stations	0.70
5 new stations	0.67

Table 5
Click here to download Table: Table 5.docx

Table 5. Optimal solutions for the five scenarios and corresponding interdistances

Increasing the existing network by	Code of the selected candidate(s)	Minimum interdistance (km)	Maximum interdistance (km)	Mean interdistance (km)	Mean drainage area (km^2)
Initial network		10.9	168.5	74.4	4246
1 new station	{C14}	10.9	168.5	73.1	3958
2 new stations	{C14, C4}	10.9	171.7	78.8	3718
3 new stations	{C14, C4, C5}	10.9	171.7	81.0	3484
4 new stations	{C14, C4, C5, C7}	10.9	171.7	81.4	3284
5 new stations	{C14, C4, C5, C7, C15}	10.9	171.7	79.4	3105

Figure captions

Figure 1. Gauged basin outlets locations for the 19 stations

Figure 2. Basins centroids and oulets of "grid" nodes for M=20 selected basins
Figure 3. Centroids of the fifteen candidate locations for composing the optimal network
Figure 4. The logarithms of observed average specific annual module versus the logarithms of drainage area
Figure 5a. Linear regression relation between the logarithm of scaled specific discharge Q_N and the logarithm of drainage area A.

Figure 5b. Linear regression relation between the logarithm of scaled specific discharge Q_N and the logarithm of basin mean rainfall P (for 21 gauged basins with rainfall information)

Figure 5c. Regression residuals versus logarithm of drainage areas

Figure 6. Calibration of the semivariogram of residuals of estimation of the scaled average specific discharge (with the corresponding sample size of pairs)

Figure 7. Spatial distribution of the optimized hydrometric network for Scenario 5 with 5 new sites.

Figure 8. The progression in mean and maximum centroids interdistance according to the Scenario (1) to (5)