Optimum Rain-Gauges Network Design of Some Cities in Iraq

Alaa MuhsainAtiaa Al-Abadi
Geology Department, Science College, Basra University
alaaatiaa@gmail.com
Ali Hassan Duhaim Al-Aboodi
Civil Engineering Department, Engineering College, Basra University
alihassan7709@yahoo.com

Abstract

A simple methodology is adapted to optimum rain-gauges network design in central and south of Iraq through simple statistical concepts, i.e, mean and coefficient of variation. Results indicate that 17 stations are required to optimize rain-gauges network with allowable error of about 10%. To redistribute the additional stations (8 stations) among the available already installation stations (9 stations) a map of standard error map of monthly average rainfall estimation using ordinary kriging technique in Geostatistical tools in ArcGIS 9.3 is used. A new optimum network is designed by the trial and error procedure via add stations to the zones with higher standard error or drop or relocate in the zone of low standard error. The methodology is simple and low cost and can be adopted to deign optimum network design over the whole area of Iraq or for other meteorological, hydrological, and hydrogeological variables.
Key words: Iraq, ArcGIS 9.3, Rain gauges, Ordinary kriging

الخلاصة:

صممت شبكة مثلى لتوزيع محطات قياس الامطار لمنطقة وسط وجنوب العراق باستخدام المفاهيم الاحصائية البسيطة (المعدل ومعامل التباين). بينت نتائج تطبيق الطريقة بان المنطقة تحتاج الى ١٧ محطة لقياس الامطار مع نسبة خطأ في معدل الامطار المقاسة بمقدار ١٠%. وبما ان المنطقة تحتوي على ٩ محطات قياس منصوبة اصلاً فان المحطات الثمانية الباقية وزعت بالاعتماد على المساحات التي تشغلها خطوط تساوي الامطار وعلى مقدار الخطأ القياسي لتخمين معدلات الامطار المحسوبة باستخدام تقنية الاستكمال الداخلي Ordinary Kriging في برنامج الــ ArcGIS 9.3. وصممت الشبكة الجديدة باستخدام طريقة المحاولة والخطأ من خلال أضافة او حذف او اعادة توزيع المحطات في المنطقة اعتماداً على مقدار الخطأ القياسي. ان الطريقة المعتمدة في التصميم هي طريقة بسيطة وقليلة الكلفة ويمكن اعتمادها لتصميم شبكة قياس امطار العراق ككل.
الكلمات المفتاحية:العراق، برنامج الــ ArcGIS 9.3، المحطات المطرية، تقنية الاستكمال الداخلي كرايج التقليدية

Introduction

The term '**precipitation**' defines all forms of water that may reach the ground surface from the atmosphere. The usual form of precipitation in south and central of Iraq is rainfall. The magnitude of rainfall varies with time and space. Its variation is responsible for many water problems such as floods and droughts. Rainfall is expressed in terms of the depth to which rainfall water would stand on an area if all the rain were collected on it (Subramany, 1994). Since the catchment area of a rain-gauge is very small compared to the areal extent of a storm, it is obvious that to get a representative picture of a storm over a catchment the number of rain-gauges should be as large as possible, i.e., the catchment area per gauge should be small. Economic, geological, accessibility may restrict the number of gauges to be maintained within an area. Hence, one objectives at an optimum density of gauges from which reasonable accurate information about the storm can be obtained.

The aim of this research is to design optimum rain-gauge network for central and south of Iraq through statistical methods and geographic information system. This study may represent a first step to build optimum rain-gauge for the whole country if sufficient data are obtained.

Methodology

The aim of the optimum rain-gauge network design is to obtain data that define statistical distribution of the meteorological elements with enough accuracy for practical situations (Das, 2002). If there is already some rain-gauge stations in an area, the optimum number of rain-gauge stations to be established in a given area is given by the following equation:

$$N = \left(\frac{C_v}{\varepsilon}\right)^2 \quad \textbf{(1)}$$

whereN = optimum number of rain-gauge stations to be installed in the area, C_v = coefficient of variation of the rainfall of the existing rain gauge stations (say *n*), and ε = allowable degree of error in the estimate of the mean rainfall (usually taken as 10%).

The number of additional rain-gauge stations (*N-n*) should be distributed in the different zones in proportion to their areas, i.e., depending upon the spatial distribution of the existing rain-gauge stations and the variability of the rainfall over the area (Raghunath, 2006). The problem is how to distribute the selected rain-gauge stations within the area itself. The solution is to use the standard error of estimate as criteria. With the help of geostatistical tool in ArcGIS 9.3 software, this work is easy to implement.

The study area and climatic data

The study area represents the south and central of Iraq between (32°59'13.71"-29° 6'32.92") latitude and(43°58'38.24"- 48°30'30.67") longitude (Fig. 1). It is encompasses area of about 142788.22 km^2. The climate of central and south of Iraq is characterized by hot, dry summer, cold winter and a pleasant spring and fall. The low latitude, the western air currents and the Arabian Gulf generally influence the climate of southern Iraq.Relatively long-term of rainfall data from Al Amarah, Al Basrah, Al Kut, Al Nasiriyah, Al Bosayyah, An najaf, Al Samawah, Karbala and stations are used as primary source of climatic information. Boston station (in Iran) is used as secondary data for facilitating extrapolation of the analysis and creating maps by using ArcGIS 9.3 software. Table (1) shows the basic information about the stations used in the study for the period (1970-2012) (meteosism.gov.iq/ar). The current rainfall network consists of nine stations. The rainfall regime of the study area is characterized by a long hot dry season from June to September where no precipitation occurs and a rainy period which is extended from October until May. The amount of rainfall in May and October is rather low. The average annual rainfall for the stations considered in the study is shown in Table (2). The seasonal rainfall averages for the stations considered in this study is shown in Table (3).

Interpolation of rainfall data using Ordinary kriging technique

Interpolation referrers to the process of estimating the unknown data values of an attribute variable for specific location using the know data values for other points. Most interpolator methods can be divided into main types namely global and local. In global method, all the variable data are used to give estimates for the points with unknown values; local interpolation methods use only the information in the neighborhood of the point being estimated. In addition, there are two groups of interpolation methods deterministic and stochastic. Deterministic methods such as inverse distance, spline, radial basis function, etc furnish no indication of the extent of possible errors, whereas stochastic methods (kriging) quantify the spatial autocorrelation among measured points and explain the spatial configuration of the sample pointsaround the prediction location. In general, all spatial interpolation methods can be represented as weighted average of sample data. They all share the same general estimation mathematical formula which is written as: (Webster and Oliver, 2001)

$$\hat{z}(x_\circ) = \sum_{i=1}^{n} \lambda_i z(x_i) \tag{2}$$

where $\hat{z}$: is the estimated value at the point of interest $x_\circ$.

z: is the observed value at the sampled point x_i.

λ_i: is the weight assigned to the sample point.

n: is the number of sample points used for the estimation

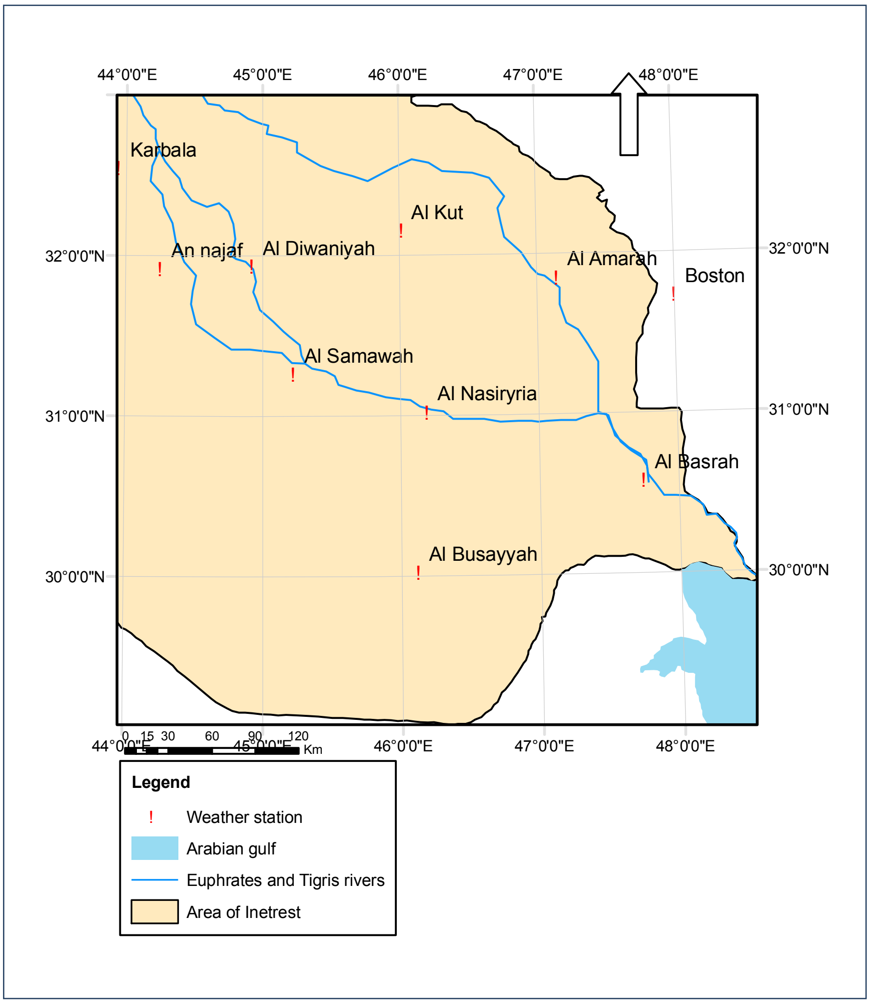


Fig. (1): location of the study area with meteorological stations.

Table (1): List of meteorological stations used in the study

Station name	Easting	Northing	Elevation(m above msl)
Al Amarah	707446.08	3523715.01	9.00
Al Basrah	766285.14	3385571.15	2.40
Al Kut	598721.75	3556971.87	149
Al Nasiyriah	616629.28	3430339.99	3.00
Boston	788030.00	3510923.92	7.80
Al Diwaniyah	493834.74	3530110.56	20.40
Al Samawah	521975.15	3457201.30	6.00
Al Najaf	432437.47	3530110.56	32.00
Karbala	405576.16	3599182.49	29.00
Al Busayyah	610233.74	3321615.66	144

Table (2): Annual rainfall at stations within the study area.

Station Name	annually total rainfall (mm)		
	Average	Max	Min
Al Amarah	172	328.2	60.1
Al Basrah	142.3	296.6	48.3
Al Kut	132.4	237.7	44.6
Al Nasiyria	109.22	221.0	25.4
Boston	165.5	396.7	28.3
Al Diwaniyah	108	223.4	37.8
Al Samawah	94.6	228.3	20.4
An najaf	68.54	167.64	33.02
Karbala	55.6	-	-
Al Busayyah	42.3	-	-

Table (3): Mean seasonal rainfall (mm) at different stations in the study area

Stations	Season			
	winter	spring	summer	Autumn
Al Amarah	32.59	16.76	0.17	12.90
Al Basrah	28.07	12.57	0.10	10.05
Al Kut	24.71	14.48	0.02	9.19
Al Nasiyria	18.53	13.55	0.00	8.89
Boston	28.27	17.27	0.50	14.25
Al Diwaniyah	19.77	10.72	0.13	8.91
Al Samawah	17.88	9.17	0.05	7.68
An najaf	16.09	10.16	0.00	7.62
Karbala	13.80	9.24	0.00	3.95
Al Busayyah	17.27	8.18	0.00	5.44

Kriging is a group of geostatistical techniques to interpolate the value of a random filed (e.g. concentrations of chemical constituent, groundwater level, land elevation … etc as a function of the geographic locations) at an unobserved location from observation of neighboring values. The kriging method is named in honor of D. G. Krige, a mining engineer in the gold fields of South Africa. The term was coined originally as Krigeage by P. Carlier, but Matheron(1963) brought it into the English language in recognition of Krige's contribution to improving the precision of estimating concentration of gold and other metals in ore bodies and recoverable resources. Kriging, like most interpolation technique, is built on the assumption that things that are close to one another are more alike than those farther away (spatial autocorrelation). The detail description of kriging theory and its application is beyond this text and could be found it in Delhomme (1978), only a brief description of the method is illustrated. Four steps are required to build ordinary kriging model. The first step is to construct a variogram from scatter observation points set to be interpolated. The variogram characterizes the spatial continuity or roughness of a data set. Variogram analysis consists of the experimental variogram calculated from the data and the variogram model fitted to the data. Suppose that the value to be interpolated is referred to as z. the experimental variogram is found by calculating the variance σ of each point in the set with respect to each of the other points and plotting the variance versus distance between the points. Several formulas can be used to compute the variance, but it is calculated by averaging one – separation distance and direction. It is plotted as a two – dimension graph. The variogram model is chosen form a set of mathematical functions that describing spatial relationship. The second step is fit a model. This is done by defining a line that provides the best fit through the points in the experimental variogram cloud graph.A model variogram is just a simple mathematical function that models the trend in the experimental variogram. In the third step the constructed model is used to compute the weights in kriging simultaneous equations. The fourth step is to make a prediction.

Modeling the semivariogram (experimental variogram)

A semivariogram is one of the significant functions to indicate spatial correlation in observations measured at sample locations. It is commonly represented as a graph that shows the difference in measure with distance of all pairs of sampled locations. In spatial modeling of the semivariogram, begin with a graph of the empirical variogram, computed as: (ESRI, 2001)

$$\gamma(x_i, x_\circ) = \gamma(h) = \frac{1}{2}\operatorname{var}[Z(x_i) - Z(x_\circ)] \qquad (3)$$

where h is the distance between point x_i and $x_\circ$ and $\gamma(h)$ is the seimvariogram (Webster and Oliver, 2001; ESRI, 2001). A state previously, a plot of $\hat{\gamma}(h)$ versus h is known as the experimental variogram (Fig. 2), which displays several important features (Burrough and McDonnel, 1998). As seen from this graph, at a certain distance the model levels out. The distance (on x-axis) when the model first flatted out is known as the range (a).Sample locations separated by distance closer than the range are spatially autocorrelated, whereas locations farther apart than the range are not. If at a distance nearly equal to zero, i.e, h→0, the variogram value is greater than zero, this value is known as the 'nugget effect' C_0. The value that the semivariogram model attains at the range (the value on the y-axis) is called the sill (*S*).The partial sill (*C*) is the sill minus the nugget. Both C_0 and the sill (*S*) characterize the random aspects of the data, where the range (*a*) and *C* characterize structural aspect of the phenomena under consideration. The nugget affect can be attributed to measurement errors or spatial source of variation at distance smaller than the sampling interval (or both).These errors occur because of the error inherent in measuring devices.

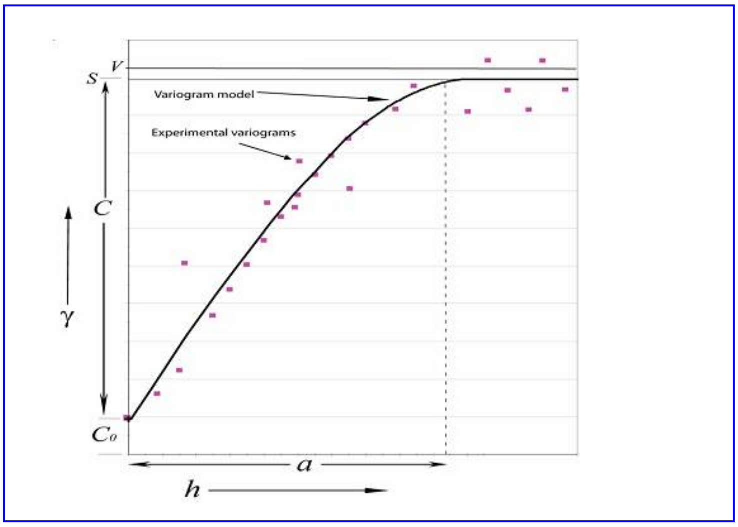


Fig. (2): A spherical variogram model fitted to the experimental variogram points (by Burrough and McDonnel, 1998).

The semivariance can be estimated from the data as follows:(Burrough and McDonnel, 1998).

$$\hat{\gamma}(h)=\frac{1}{2n}\sum_{i=1}^{n}\left(z(x_i)-z(x_i+h)\right)^2 \qquad (4)$$

where$n$ is the number of pairs of sample points separated by distance h.

Once the semivariogram is created, the next step is to fit mathematical function (e.g. line) to the points of semivariogram cloud graph. The modeling of semivariogram is similar to fitting a least – square line in regression analysis. Some of the useful models are available including linear, spherical, experimental, Gaussian, etc, Fig. (**3**)and Table(**4**).

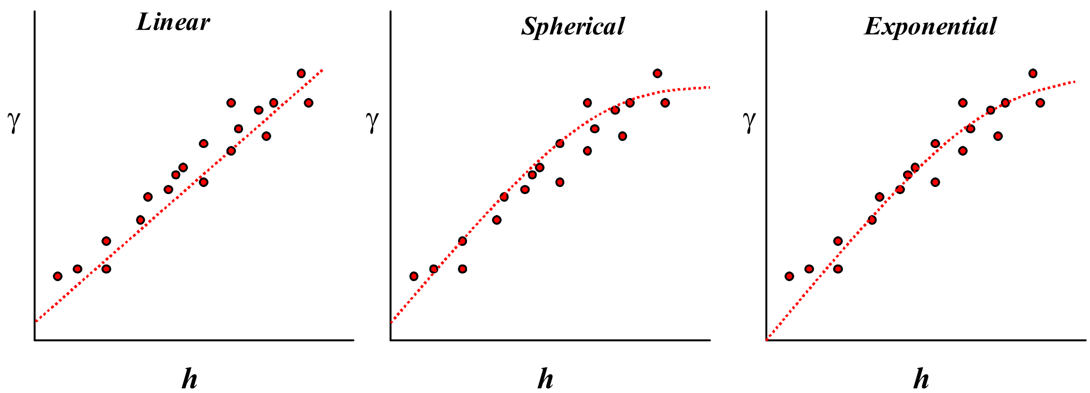


Fig. (3): Different types of variogram models (modified after Burrough and McDonnel, 1998)

Table (4): Mathematical equations of main types of variogram models (modified after Yang *et al*. (2008))

Variogram model type	Variogram $\gamma(h)$ mathematical formula
Exponential	$\gamma(d) = C_0 + C\left(1 - e^{-\frac{d}{\alpha}}\right)$
Gaussian	$\gamma(d) = C_0 + C\left(1 - e^{-\frac{d^2}{\beta^2}}\right)$
Spherical	$\gamma(d) = C_0 + C\left(\frac{3}{2}\frac{d}{\lambda} - \frac{1}{2}\left(\frac{d}{\lambda}\right)^3\right)$

C_0 is the nugget effect, $C_0 + C$ or λ is the sill, *d* is the distance, α and β are the length parameters.

Kriging Estimator

All kriging estimator are variants of the basic equation (3), which is a slight modification of equation (2), as follows: (Wackernagel, 2003)

$$\hat{Z}(x_\circ) - \mu = \sum_{i=1}^{n} \lambda_i [Z(x_i) - \mu(x_\circ)] \quad (5)$$

where μ is a known stationary mean, assumed to be constant over the whole domain and calculated as the average of the data. The parameter λ_i is kriging weight; n is the number of sampled points used to make the estimation and depends on the size of the search window; and $\mu(x_\circ)$ is the mean of samples within the search window. The kriging weights are obtained by minimizing the variance as: (Isaaks and Srivastava, 1989)

$$\text{var}\left[\hat{Z}(x_\circ)\right] = E\left[\left\{\hat{Z}(x_\circ) - Z(x_\circ)\right\}^2\right]$$

$$= E\left[\left(\hat{Z}(x_\circ)\right)^2 + \left(Z(x_\circ)\right)^2 - 2\hat{Z}(x_\circ)Z(x_\circ)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(x_i - x_j) + C(x_\circ - x_\circ) - 2\sum_{i=1}^{n} \lambda_i C(x_i - x_\circ) \quad (6)$$

where $\hat{Z}(x_\circ)$ is the true value expected at point $x_\circ$, *n* represents the number of observations to be included in the estimation, and $C(x_i - x_j) = Cov\left[Z(x_i), Z(x_j)\right]$.

In ordinary kriging (OK) the value of the attribute is estimated using equations 4 and 5 by replacing μ with local mean $\mu(x_\circ)$ that is the mean of samples within the search window and forcing $\left[1 - \sum_{i=1}^{n} \lambda_i\right] = 0$, that is $\sum_{i=1}^{n} \lambda_i = 1$, which is achieved by putting it into equation 5 (Clark and Harper, 2001).

Many authors attempt to use kriging methods for precipitation mapping on daily basis (e.g. Carrera-Hernandez and Gaskin, 2007; Kyriakidis *et al*., 2001; Symeonakis *et al*., 2009), monthly (e.g. Lloyd, 2005) and annual averages (e.g. Hofierka *et al*., 2002; Goovaerts, 2000; Martínez-Cob, 1996), and satisfactory results were obtained. Because of many advantages of this interpolator method, it is used here for interpolating annual rainfall average which is explained in the next section.

Data exploratory

To interpolate annual rainfall average for the study area using ordinary kriging technique, Geostatistical extension of ArcGIS 9.3 is used. Before building different variogram models for data set, Exploratory Spatial Data Analysis (ESDA) is performed. The kriging interpolation technique gives the best results if the data is normally distributed (bell – shaped curve). Therefore, it is important to understand the data before creating a surface. ESDA allows examining a given data in different ways and enables to gain deeper understanding of the data that investigate so that a better decision can make from it. After finishing the ESDA, Three main different variogram models (Spherical, Exponential, and Gaussian) for observed rainfall levels were plotted and the semivariogram parameters (i.e., range, sill, and nugget) were obtained with cross-validation results, Table (**5**). Cross-validation gives an idea of 'how well' the model predicts the unknown values. This test helps to make a decision about which model provides accurate prediction. The calculated error statistics serve as diagnostic that indicates whether the model is reasonable for map prediction. In this study two error statistics are adopted: the mean error (m) and the root-mean-square standardized error (η). The model will be reasonable for prediction if the mean error approaches zero and the root-mean-square standardized error approaches 1. The two error statistics are computed as follows: let $\hat{Z}(x_i)$ is the predicted value using interpolation method, let $Z(x_i)$ is the observed value at the x_i location, and let $\hat{\sigma}(x_i)$is the predicted standard error for the location $x_\circ$. Then the mean prediction errors (m) is

$$m = \frac{\sum_{i=1}^{n}\left(\hat{Z}(x_\circ) - Z(x_\circ)\right)}{n} \qquad (7)$$

The root-mean-square standardized error (η) is calculated as

$$\eta = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(\hat{Z}(x_\circ) - Z(x_\circ)\right)/\hat{\sigma}(x_\circ)\right]^2}{n}} \qquad (8)$$

where n is the number of values in dataset.

As shown in Table **(5)** the three variogram models have the same nugget, range and approximately same values of sill. Because the exponential model has the lowest error statistics, the model is chosen for the interpolating annual rainfall average over the study area, Fig. **(4).** This map clearly shows that the annual rainfall average increases from southwest to northeast.

Table (5):Variogram parameters for fitted experimental dataset and cross-validation results

Model variogram	Sill	Range	Nugget	m	η
Spherical	2586.5	378641	0	3.71	0.67
Exponential	3251.7	378641	0	2.49	0.72
Gaussian	3581.3	378641	0	5.75	2.04

Results and discussion

The optimum number of rain-gauge stations to limits the error in the mean value of rainfall to be ε = 10%, hence

$$N = \left(\frac{C_v}{\varepsilon}\right)^2 = \left(\frac{41.0}{10}\right)^2 = 16.8 \approx 17$$

Therefore, the required additional stations is (17-9) = 8 stations. Table **(6)** shows how to distribute these stations according to the area between two isohyetals.

Table (6): Additional stations required to optimize rainfall network

Zone	I	II	III	IV	V	VI	Total
Area (km^2)	50877	10488	22560	25089	20282	13492	142788
Area as decimal	0.35	0.07	0.16	0.18	0.14	0.09	1
N ´ Area in decimal (N=17)	6.06	1.25	2.69	2.99	2.41	1.61	
Rounded as	6	1	3	3	2	2	17
Rain-gauges existing	2	1	3	1	1	1	9
Additional rain gauges	4	-	-	2	1	1	8

To optimize location of additional stations, a mean standard error of kriging error is used as a criterion. The smaller the variance of the estimation error is the more effective can be the monitoring serve the objective. The standard error map of the estimation is shown in Fig. **(٥)**. From this figure with the required number of observation stations from network density, new stations are added in the zones with higher standard error while some are dropped or relocated in the zone of low standard error. A new monitoring network for this study area is designed, Fig.**(٦)**, according to this procedure. The other factors which may be considered are the accessibility of rain-gauges locations and the cost of installation, hence, a field campaign must be organized to achieve this goal.

Conlcuions

To optimize rain-gauges network in central and south of Iraq, eight (8) additional stations must be installed in addition to the pre-installed stations. The number of additional rain-gauges is calculated using a simple statistical formula through mean and coefficient of variation. Distribution of these stations is accomplished through trial and error procedure depending on the standard error of estimation as a criterion. The adapted method is simple, low cost, and robust and the results can be generalized to design an optimum rain-gauge network for the whole of country.

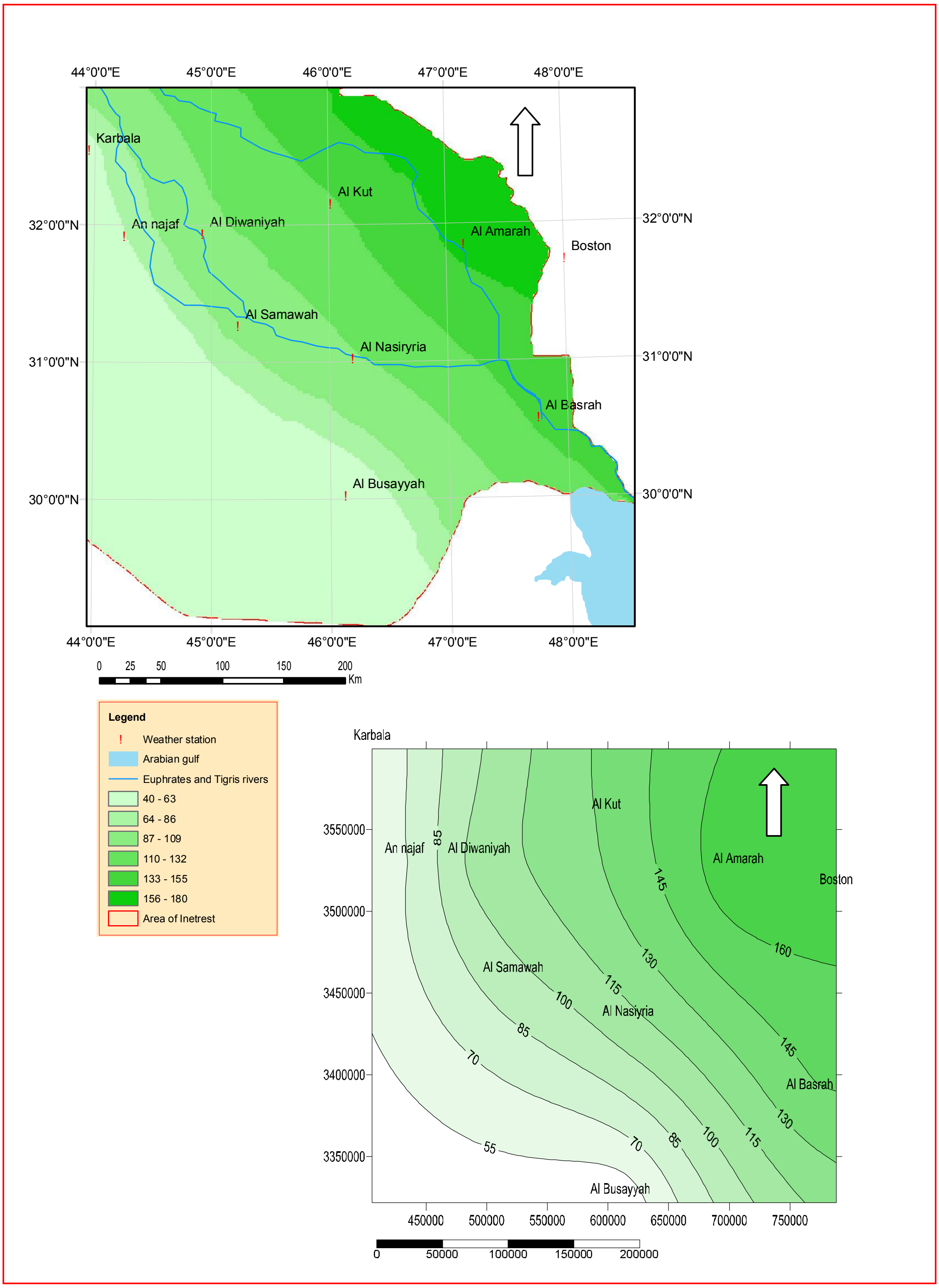


Fig. (4): Spatial distribution of average of total annual rainfall (isohyets) in the study

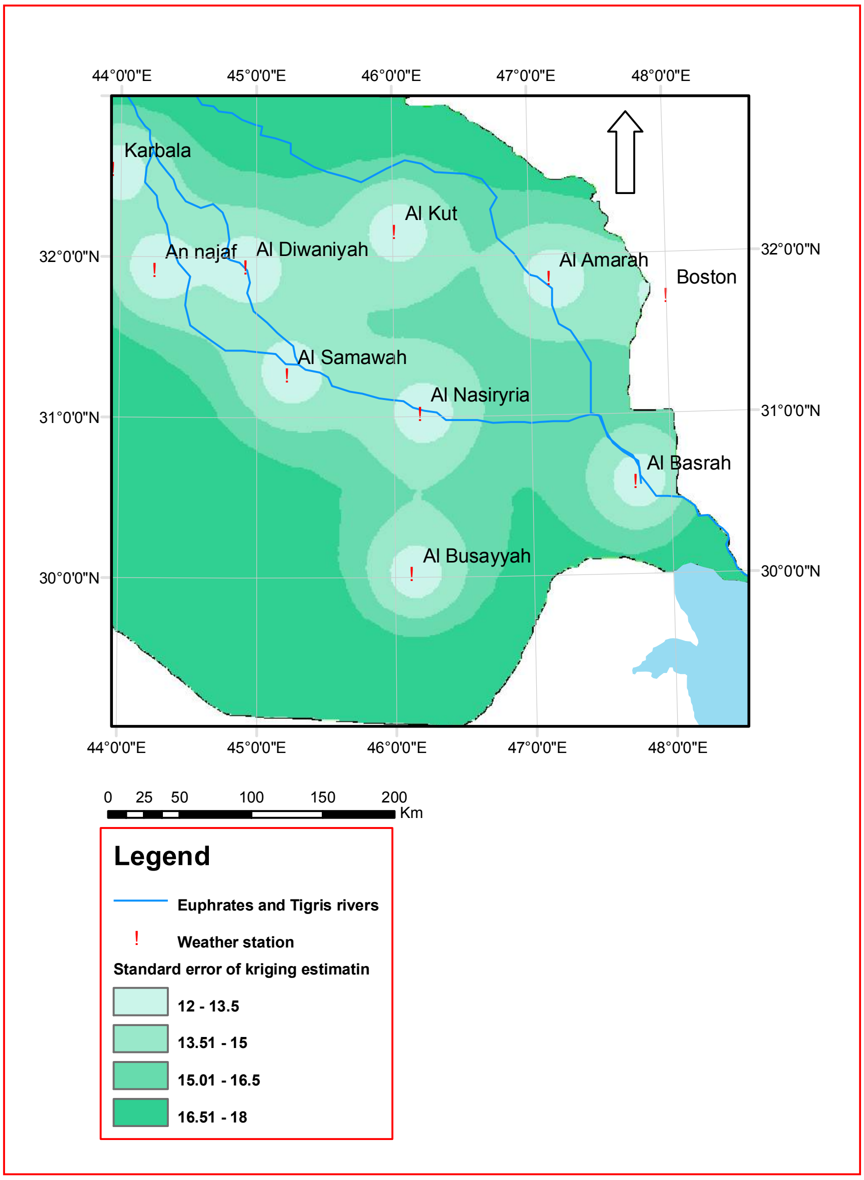
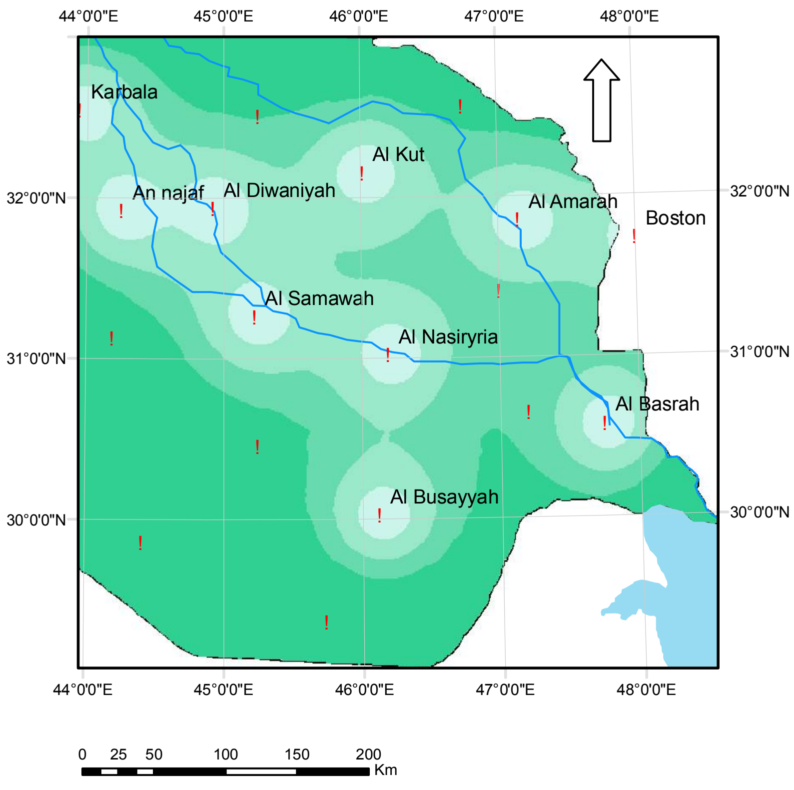


Fig. (5): Standard error map of kriging estimation.



Fig(6): Locations of additional required stations to optimize rain-gauge network

References

Burrough, P.A. and McDonnell, R.A.,1998. Principles of Geographical Information Systems. Oxford University Press, Oxford, 333 pp.

Carrera-Hernandez, J.J., Gaskin, S.J., 2007. Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico. Journal of Hydrology 336,231–249

Clark, I. and Harper, W.V.,2001.Practical Geostatistics 2000.Geostokos (Ecosse) Limited, 342pp.

Das, G., 2002. Hydrology and Soil Conservation Engineeing, Prenctic-Hall of India, New Delhi, 489p.

Delhomme, J.P.,1978. Kriging in the hydrosciences. Advances in Water Resources,1(**5**): 251-266.

ESRI, 2001.Using ArcGIS Geostatistical Analyst, USA, 306p.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into thespatial interpolation of rainfall. Journal of Hydrology 228, 113–129.

Hofierka, J., Parajka, J., Mitasova, H., Mitas, L., 2002. Multivariate interpolationofprecipitation using regularized spline with tension. Transactions in GIS 6 (2),135–150.

Isaaks, E.H. and Srivastava, R.M., 1989. Applied Geostatistics. Oxford University Press, New York, 561 pp.

Kyriakidis, P.C., Kim, J., Miller, N.L., 2001. Geostatistical mapping of precipitationfrom rain gauge data using atmospheric and terrain characteristics.Journal ofApplied Meteorology 40, 855–1877.
Lloyd, C.D., 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. Journal of Hydrology 308,128–150.
Martínez-Cob, A., 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. Journal of Hydrology 174,9–35.
Matheron, G., 1963. Principles of Geostatistics. Economic Geology J. 58**:** 1246-1266.
Raghunath, H. M., 2006. Hydrology: Principle, analysis, and design. New Age International Publishers, New Delhi, India. 477pp.
Subramanya, K.,1994. Engineering Hydrology. Tata McGraw-Hill Publishing Company L., New Delhi. 392pp.
Symeonakis, E., Bonifacio, R., Drake, N., 2009. A comparison of rainfall estimationtechniques for sub-Saharan Africa. International Journal of Applied Earth Observation and Geoinformation 11, 15–26.
Wackernagel, H.,2003. Multivariate Geostatistics: An Introduction with Applications. Springer, Berlin, 387 pp.
Webster, R. and Oliver, M., 2001.Geostatistics for Environmental Scientists. John Wiley & Sons, Ltd, Chichester, 271 pp.