



Assessment of groundwater contamination by different interpolation methods for water resources management in the Mitidja Plain aquifer (North-Center Algeria)

Ouahiba Aziez^{a,c,*}, Boualem Remini^b, Mohammed Habi^c, Abdelhadi Ammari^a

^aWater Engineering and Environment Laboratory, Higher National School of Hydraulics, BP 31 Soumaa Blida, Algeria, Tel. +213 560 339 160, email: o.aziez@ensh.dz (O. Aziez), Tel. +213 550 578 141, email: a.ammari@ensh.dz (A. Ammari)

^bDepartment of Water and Environmental Sciences, Faculty of Technology, Blida1 University, Algeria, Tel. +213 665 607 486, email: reminib@yahoo.fr (B. Remini)

^cDepartment of Hydraulics, Faculty of Technology, University of Tlemcen, Algeria, Tel. +213 551 278 443, email: moha.habi@gmx.de (M. Habi)

Received 20 October 2017; Accepted 8 August 2018

ABSTRACT

Due to a reported case of groundwater contamination, the present research is conducted to determine the degree of groundwater contamination present in the plain of Mitidja, North-Algeria. Selecting a suitable interpolation method to produce Piezometric maps consisting of averages sampled from 34 wells. Achieved by analyzing the effects of four spatial interpolation methods on groundwater contamination, including Empirical Bayesian Kriging (EBK), Inverse Distance Weighting (IDW), Ordinary Co-Kriging (OCK) and Ordinary Kriging (OK), with regard to the water quality index (WQI). These methods are widely used by applying numerical values to establish a range of groundwater quality data points and map the contaminated areas. They are crucial decision support tools used by managers to assess groundwater resource potential and for general management functions. The datasets used where collect from 14 aquifers across the plain of Mitidja. The evaluation is used to model the groundwater contamination areas, where the spatial uncertainty of the contaminated areas appears prominently between the transition level of one contaminated area to another. Also, cross-validation and various contaminated surface areas are used to assess the performance of each interpolation. The outcomes indicated that the performance differed slightly among different methods. The subtraction results showed a clear spatial difference amongst the contamination assessment results. Results of both the Ordinary Co-Kriging (OCK) and Empirical Bayesian Kriging (EBK) methods showed to have minor differences and the weakest RMSE values. The Inverse Distance Weighting (IDW) showed a healthy relationship between the weighting power of IDW and the groundwater contamination.

Keywords: Water quality index (WQI); Groundwater contamination; Spatial interpolation; Inverse distance weighting; Empirical Bayesian Kriging; Algeria

1. Introduction

The human body holds up to 60% of water, the brain consists of 73% water, the lungs 83% water and even the bones are watery holding up to 31% water. The consumption of chemically contaminated water contributes to a range of health issues who depends on the toxicity of the dissolved chemical elements. The international water qual-

ity standards are used, to estimate water quality and to confirm several quality indices. The vast plain of Mitidja is part of a lively agricultural area. Its groundwater resources are used for drinking water, irrigation and industrial activities.

Evaluating groundwater contamination and its spatial distribution are especially significant for health risk assessment [1]. Groundwater management in this aquifer involves the monitoring of water quality across the water points located on the surface. The number of sample data is important, in determining precision. In the general sur-

*Corresponding author.

This article was originally published with an error in an author's name and email. This version has been corrected. Please see Corrigendum in vol. 172 (2019) 428 [10.5004/dwt.2019.25315].

veillance networks of the wells are not representative, due to excessive cost constraint [2], and only a small percentage of sample data can be analyzed, to determine the quality and quantity of groundwater in use [3]. Therefore, the data measured contains a considerable amount of uncertainty [4]. The interpolation methods are needed to determine spatial variety of groundwater levels [2] and are powerful tools to be used to predict surface values [5]. It is possible to produce maps using the Kriging method which delivers the best results and improve the qualitative and quantitative management of water resources [6]. Geostatistics initially used in the mining sector by Krige, is used to monitor groundwater salinity, to map contaminated areas affected by heavy metals, nitrates evolution and fluctuation in groundwater levels, for example [2,3,7–11].

Two branches arise from regionalization: Determinist and Geostatistical, the first, produces a surface from a measured function point to similarity spreading, for example (IDW) [2,3,7–11], the second, are based on statistical properties of measured levels, for example Ordinary Co-Kriging (OCK) and the Kriging Bayesian Empiric (EBK). We cannot validate that one method is better than another, this depends on work results. Amongst researchers, the Kriging's method gives the best results [12–16], while other sources showed that IDW gives the best result [5,17]. Other authors found satisfactory results in using the Kriging Bayesian Empiric (KBE) [3,18–20]. Interpolation methods all have a smoothing effect, which underestimates high local values and overestimates local weak values [3,21,22]. This smoothing effect leads to a bias in contamination evaluation and has an impact on the pertinent environmental decision making [3,22,23]. Ordinary Kriging (OK), Inverse Distance Weighting (IDW) and Ordinary Co-Kriging (OCK) are the most frequently compared methods [24,25]. Ordinary Co-Kriging (OCK) advances over Kriging only when the secondary variables are better sampled than the primary variable, or more accurately reflect the real world [25]. Consequently, interpolation methods such as IDW and OK have been extensively used

in groundwater investigations and contamination mapping [3,5,8]. Likewise, there are other studies in which the Empirics Bayesian Kriging (EBK) interpolation method has outperformed other methods such as [3,19].

Therefore, this study aims to show spatial-temporal comparisons between four interpolation methods; Ordinary Kriging (OK), the Kriging Bayesian Empiric (EBK) and Inverse Distance Weighted (IDW) and so Ordinary Co-Kriging (OCK) between Water Quality Indices (WQI) and Piezometric groundwater levels to determine the best estimation method to determine groundwater quality in the plain of Mitidja.

2. Material and methods

2.1. Study area

It is important to carrying out a preliminary study on hydrochemistry and the possible contaminated sources, where groundwater quality is affected by geological formations and by anthropogenic activities [26,27]. Agricultural and industrial activities are significant factors that contribute towards groundwater contamination. The chemical compound of groundwater is controlled by, rainfall composition, geological structure, and aquifers mineralogy [28]. The groundwater exhaustion rate is one of the deterioration factors of groundwater quality [29,30]. The interaction of all these factors leads to various water facies [31,32].

2.1.1. Geographic location

The Mitidja is a vast plain situated in North-Algeria that stretches across 1450 km² south of the Sahel of Algiers region. The plain is found between (36°29'N to 36°44'N) North latitude and (2°25'E to 3°17'E) East longitude. It extends over four provinces (Wilaya): Algiers, Blida, Tipaza, and Boumerdes. It is located on an axis of subsidence-oriented ENE-WSW. (Fig. 1).

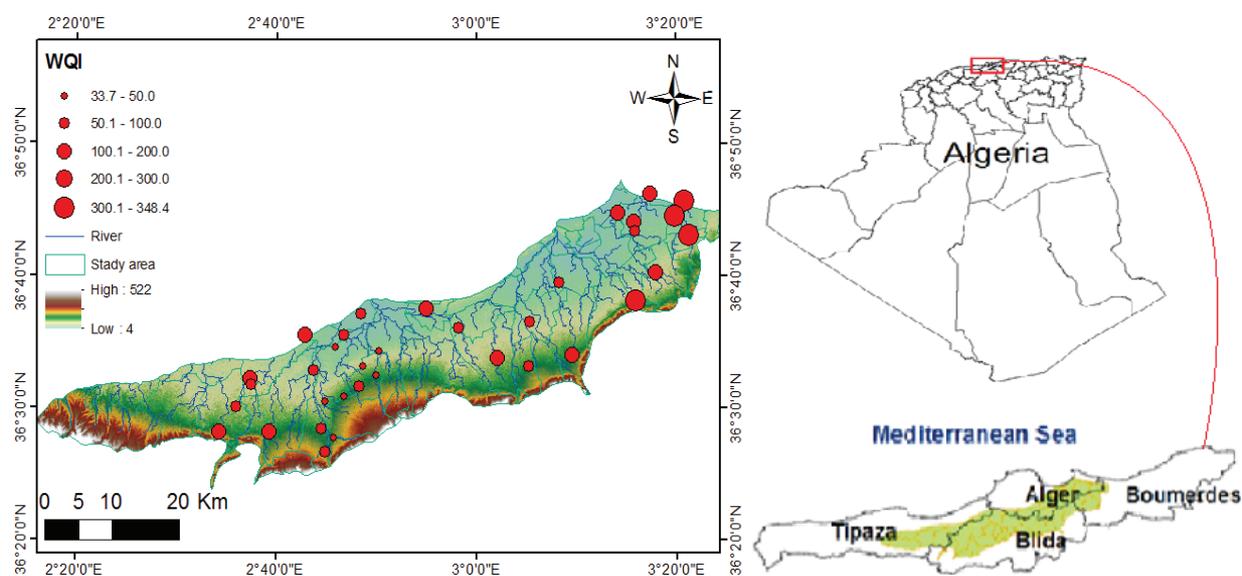


Fig. 1. Location map of Mitidja plain showing sampling sites.

2.1.2. Geology and hydrology

The Mitidja is sub-divided into three large geological areas: The Atlas Blideen which is a mountainous massif formed by joined sedimentary formation, fluvial siliceous gravel, and sandstone gravel with the red clay of Cretaceous origin; The Sahel south piedmont; Mio-Plio-Quaternary deposits consists of clay, Marls and limestones (aquitar). The Astian aquifer, is the deepest confined aquifer, making it less exposed to contamination and alluvial groundwater of Mitidja (quaternary unconfined groundwater). The clay formations of El Harrach separate these regions. In the hydrographic basin of Algiers, the groundwater resources of Mitidja is of 328 hm³ and found within the main capturing fields, Mazafran I and II, Chebli, in the West, and Baraki, HaouchFelit, Hamiz in the East. The aquifer recharge is made up of the wadis: El Harrach, Chiffa, Mazafran, Hamiz, infiltration coefficient into alluvium is 10% for recent alluvium and 15% for earlier alluvium; storage coefficient is around 3% in the West and 15% in the East.

2.1.3. Piezometric map

Surf Software 9 is used to create a piezometric map which consists of averages sampled over two full water points, low water, and high water, across 35 wells, it presents a divide in groundwater SSE-N NW. Values of ISO piezometric curves decreased in the sub-triangle of cation from South to North (Fig. 2).

2.1.4. Rainfall

The study area has a Mediterranean climate, characterized by hot and dry summers (May to September), and rainy winters (October to April), with an average rainfall around 607, 25 mm, the average temperature is 18.5°C, and the annual rate of evapotranspiration is 1240 mm. (Fig. 2). Spatial rainfall statistics collected in 2010 from 14 stations surrounding the study area produced a larger rainfall gradient. Ranging from 950 mm in the Ouled El Alleug, it is at 36.55528 N 2.79028 E, to less than 500 mm in Ameur el Ain, it is at 36.5062° N, 2.5693° E and in Algiers, it is at 36.7538° N, 3.0588° E. Over 80% of these rainy episodes occurred during the cold, rainy winter months.

2.1.5. Agriculture and industry

The soil characteristics, of the agro-pedological study, of Mitidja highlighted five soil classes: poorly evolved soils, calci-magnesian soils, iron sesquioxide soils and hydromorphic soils. Hydromorphic soils can be found across all important study areas. The Mediterranean climate is favorable to different cultures: cereals, citrus, fruit trees, vegetables, vine and industrial crops, such as tobacco. The land use map showing occupied agricultural ground areas (Fig. 2). This region includes an important food reservoir able to give a high proportion of agricultural products to the population. As for industry and according to a report from the Water Resources Minister MRE [33] and Mutin G.

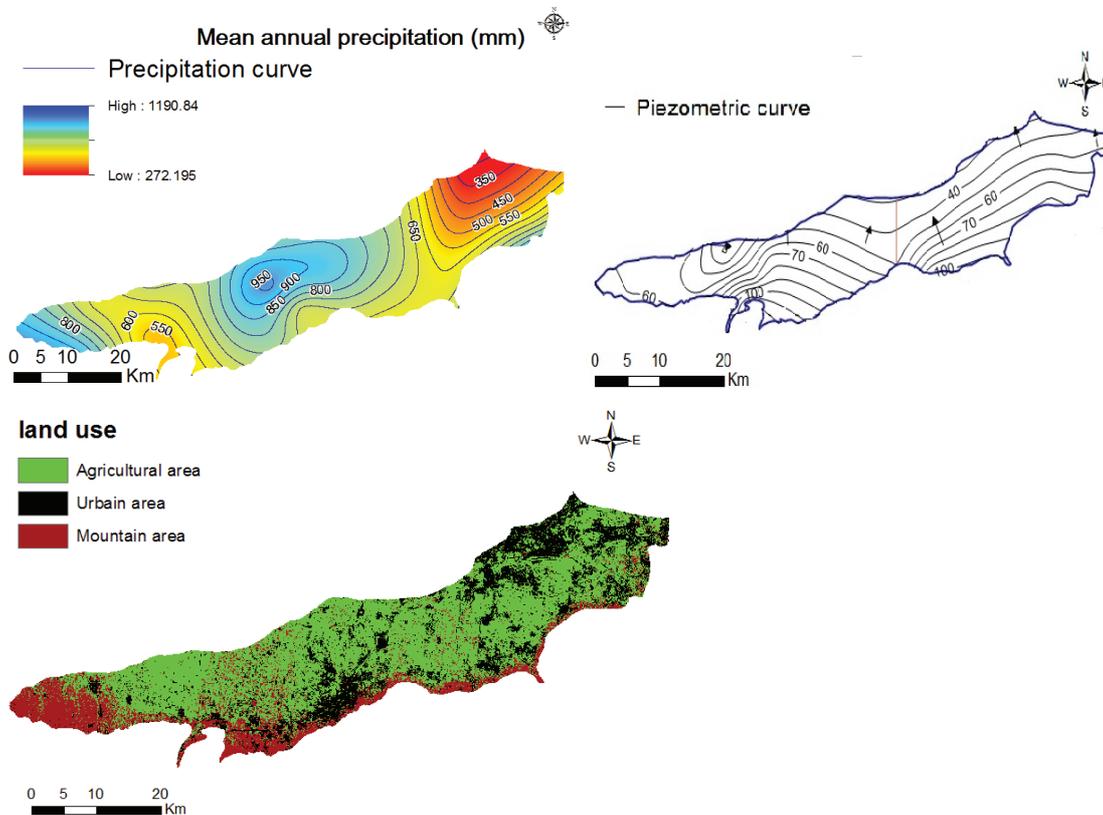


Fig. 2. Top left: mean annual precipitation (2010). Bottom left: Land use map. Top right: Piezometric level of the Mitidja (2010).

[34], major industrial activities take place in the Centre and East of the plain where the following industries are present: metal, steel and plastic manufacturing, metallurgical operations, agro-food, transformation industry, chemistry.

2.2. Hydrochemical approach

Before calculating the Water Quality Index (WQI), the hydrochemical facies data are analyzed using the Piper diagram. This diagram is produced with the use of free DIAGRAM software, developed by the Hydrogeology Laboratory of Avignon. *Hydrogeochemical facies interpretation* is a useful tool for finding the flux profile and the origin of the chemical composition of groundwater; it is used to express similarity and dissimilarity in the chemistry of groundwater samples based on dominant cation and anion of groundwater [28]. The Piper diagram, which is composed of two triangles, representing the cationic and the anionic facies and a lozenge synthesizing the global facies.

Water samples are analyzed for *physico-chemical* parameters to obtain an average of two sampling periods from the groundwater (low water and high water), across 35 water points. The analysis is completed in the Laboratory of National Agency of Water Resources (ANRH) and focused on these major elements (Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , HCO_3^- , NO_3^-) and heavy metals (Pb, Cu, Cd, Fe, Zn, Cr).

Water classification according to the Piper Global Diagram (Fig. 3) showed facies of chloride sulfate calcium and magnesium in the wells located in the East and West Zones of the plain and bicarbonate, calcium and magnesium located in the Central Zone of the plain. (East Zone: RED and YELLOW, West Zone: GREEN and BLUE, Centre Zone: PINK).

The central zone of the plain has a tendency towards the calcium pole in the sub-triangle of cations water, while

a bicarbonate trend is present in some water points in the center and west parts of the plain, far from the sea, near the sea a chloride trend is present in the east and west parts of the plain. The saline invasion of the groundwater appeared in the 1980's and aggravated in the 1990's at coastal well level. The calcium bicarbonate magnesium facies resulted from the dissolution of Astian clay, limestones, and limestone-sandstone.

3. Database and methodology

This study presents the results of 35 water samples, spread across all surfaces in the study zone with an average of two periods: low water period (dry period) and high-water period (spring period) in the year 2010. The hydrogeochemistry of Mitidja aquifer is studied covering 14 variables, Ca^{2+} , Mg^{2+} , Cl^- , SO_4^{2-} , NO_3^- , PH, Electrical Conductivity Ca, CE, TDS, TH, Fe, Mn, Cu, Pb Cd. Water samples are collected periodically by the National Agency of Water Resources (Blida and Algiers). The minimum, maximum, average, and standard deviation of these parameters are shown in (Table 1).

To verify the reliability of the results obtained, the calculation of the ionic equilibrium of the water is carried out, considering the relation between the total cations and the total anions [35]. This shows percentages ranging from -5% to $+5\%$, which corresponds to the acceptable reliability of the unit of chemical results. The total concentration of cations is calculated as the sum of calcium, magnesium, sodium, and potassium. The total concentration of anions is calculated as the sum of chloride, carbonate, bicarbonate, and sulfate. The following equation is used:

$$\text{IBE} = (\Sigma \text{Cations} - \Sigma \text{Anions}) / (\Sigma \text{Cations} + \Sigma \text{Anions}) \times 100 \quad (1)$$

where the concentration of ions is expressed in meq/L.

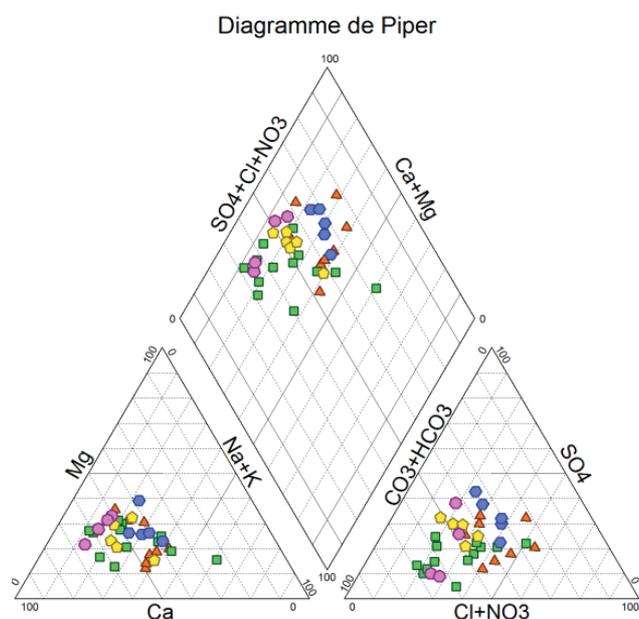


Fig. 3. Groundwater samples plotted on Piper trilinear diagrams.

Table 1
Water quality parameter, WHO Standards, and weightages

Paramètre	Standard SI	Weight (wi)	Relative weight (Wi)
pH	7.0–8.5	4	0.1000
Ca mg/l	75	1	0.0250
Mg mg/l	30	1	0.025
Cl mg/l	250	2	0.050
SO ₄ mg/l	200	2	0.050
NO ₃ mg/l	50	5	0.125
CE μS/cm	750	3	0.075
TDS mg/l	500	4	0.100
TH °F	100	2	0.050
Fe mg/l	0.1	3	0.075
Mn mg/l	0.1	3	0.075
Cu mg/l	2	2	0.050
Cd μg/l	0.003	4	0.100
Pb mg/l	0.01	4	0.100
Total		40	1.000

3.1. Calculation of the water quality index

In this study, the Water Quality index is calculated according to the indexing method suggested by Horton [36] and checked by several searchers Rupal et al. [37]; Tiwari et al. [38]; Mirzaei and Sakizadeh [3]; Sakizadehand Ahmad-pour [39]. Forth steps have been monitored by Rupal et al., [37] and Tiwari et al., [38].

3.1.1. First step

A weight (wi) is assigned to each parameter according to its relative importance in the global drinking water quality. The maximum weight of 5, for example, is assigned to the nitrate parameter due to its harmfulness on consumers' health when drinking polluted groundwater. Other parameters like calcium, magnesium, sodium, and potassium are assigned a weight between 1 and 5 depending on their importance in the overall water quality for drinking purposes [37,38]. Magnesium which is considered for the minimum weight of 1 as magnesium itself may not be harmful.

3.1.2. Second step

The relative weight (Wi) is calculated by the following equation: [3,36–38]:

$$Wi = wi / \sum_{i=1}^n wi \tag{2}$$

Wi is the relative weight, wi is the weight of each parameter and n is the number of parameters.

3.1.3. Third step

A quality scale (qi) for each parameter is assigned by dividing its concentration in each water sample by its respective norm according to the established norms by WHO (2004), and the result is multiplied by 100:

$$qi = Ci/Si \times 100 \tag{3}$$

where qi is the quality scale. Ci is the concentration of each chemical parameter in each water sample in mg/l LWHO (2004) the norm of drinking water for each chemical parameter in mg/L.

3.1.4. Fourth step

The Sli is first determined for each chemical parameter, which is then used to determine the WQI by the following equation:

$$Sli = Wi \times qi \tag{4}$$

where Sli is the sub-index of i th parameter, qi is based on the notation on the concentration of i th parameter, n is the number of parameters. The WQI is calculated by the following equation:

$$WQI = \sum Sli \tag{5}$$

The weightage values obtained by the above method are shown in (Table 1).

Based on the above WQI values, the groundwater quality is rated as excellent, good, poor, very poor, and unfit for human consumption (Table 2).

3.2. Interpolation methods

In this work, the Interpolation Methods, Inverse Distance Weighted (IDW), Ordinary Kriging (OK), Ordinary Co-Kriging (OCK) and Empirical Bayesian Kriging (EBK) have been measured.

3.2.1. Inverse distance weighting

Inverse distance weighting is based on the premise that the predictions are a linear combination of available data. In this method, the interpolating function is as follows:

$$Z(x) = \frac{\sum_{i=1}^n wi Zi}{\sum_{i=1}^n wi} \tag{6}$$

In which

$$wi = d^{-u}$$

where Z(x) is the predicted value at an interpolated point, whereas Zi is the amount at a known point. n is the total number of known points used in interpolation, di is the distance between point i and the prediction point, wi is the weight assigned to point i. Higher weighting values are assigned to those points which are closer to the interpolated point. As the distance increases, the weight decreases, and u is the weighting power that imposes the amount of weight decrease with respect to the increase in distance [3,22].

3.2.2. Ordinary kriging

Ordinary Kriging is a linear estimator meaning that the estimate of the unknown value is a linear combination of the known data values [3,22]. The aim of Ordinary Kriging is to estimate the value of a random function, z, at one or more unsampled points or over larger blocks, from more or less sparse sample data on given support, say z(x1), z(x2), z(xN), at x1, x2, xN.

This can be shown by

$$z^*(x_0) = \sum_{i=1}^n wi Z(x_i) \tag{7}$$

Table 2
Water quality index categories

WQI range	Category of water
<50	Excellent water
50–100	Good water
100–200	Poor water
200–300	Very poor water
>300	Unfit for drinking purpose

where w_i are the weights assigned to the known value of $z(x_i)$ and $z^*(x_0)$ is the estimated value. To ensure that the estimate is unbiased, weights are made to sum to 1 [22].

$$\sum_{i=1}^n w_i = 1 \quad (8)$$

3.2.3. Empirical Bayesian Kriging

Empirical Bayesian Kriging (EBK) is a geostatistical interpolation method that automates the most difficult aspects of building a valid kriging model. In addition to accounting for the uncertainty in the underlying semivariogram parameters, the other main redeeming feature of EBK is that despite the common applied Geostatistical Analyst in ArcGIS10.2, the parameters in the new developed EBK are automatically optimized through a subsetting and simulation process which is implemented by estimating a lot of semivariogram models instead of a single semivariogram [3].

The prediction in unknown locations in common kriging methods is done through calculation of semivariogram with respect to the known data locations resulting in the underestimation of the standard error of the prediction due to overlooking the uncertainty of semivariogram. On the contrary, EBK uses an intrinsic random function as the kriging model despite the other kriging methods [40].

The other main difference of EBK with that of the other kriging models is that EBK does not assume a tendency toward an overall mean; thus, there is the same chance for large deviations to get larger or get smaller. The following process is followed in EBK. (1) Using the available data, a semivariogram model is estimated. (2) Given this semivariogram, a new value is simulated at each of the input data location. (3) With respect to the simulated data, a new semivariogram model is estimated accordingly. The calculation of a weight for the latest semivariogram according to Bayes' rule is the next step in this field. The semivariogram estimated in step 1 is used to simulate a new set of values at the input location during the repetition of steps 2 and 3. A new semivariogram model and its weight are produced given the simulated data. During this process, the predictions and their respective standard errors are produced at the unsampled locations.

This process finally creates a spectrum of semivariograms [41]. There are two base distributions available in EBK with respect to the utilized multiplicative skewing normal score in this method: empirical and log empirical. Since log empirical only accept positive data values and their predictions are also positive, so it is a good option for water quality indices that are just positive scores. Due to the application of log transformation on our water quality indices, exponential model is applied to interpolate WQI in this study.

3.2.4. Co-Kriging

Co-Kriging is an extension of kriging to situations where two (or more) variables are spatially intercorrelated. For simplicity, only one co-variable is used in the following (for

a generalization see, e.g., Journel and Huijbregts in [42]). Co-Kriging is a weighted average of observed values of the primary variable z_1 (the variable of 1 immediate interest, e.g., WQI) and the co-variable z_2 . The estimated value of the primary variable at location x_0 is:

$$\hat{z}_1(x_0) = \sum_{i=1}^{N_1} \lambda_{1i} z_1(x_{1i}) + \sum_{i=1}^{N_2} \lambda_{2i} z_2(x_{2i}) \quad (9)$$

where N_1 and N_2 are the numbers of neighbors of z_1 and z_2 ; λ_{1i} and λ_{2i} are the weights associated to each sampling point. When variables $z_1 = z_2$ the system converts to kriging. The weights are chosen to minimize the co-kriging variance by solving the co-kriging system [42].

3.3. Comparison of the interpolation methods

Cross-validation and validation with an independent dataset are the most widely used methods for comparing the interpolation methods. Because the sample size is limited, cross-validation is applied in this study. Cross-validation involves consecutively removing a data point, interpolating the value from the remaining observations, and comparing the predicted value with the measured value. The mean error (ME), mean absolute error (MAE), mean relative errors (MRE), mean squared error (MSE), root mean squared errors (RMSE), Nash–Sutcliffe efficiency (NSE), and percentage bias (PBIAS) calculated from the measured and interpolated values at each sample site are used to compare the accuracy of predictions.

ME is used for determining the degree of bias in estimates and is often referred to as “bias”, RMSE and MSE provides a measure of error size, where as MAE is less sensitive to extreme values and indicates the extent to which the estimate can be in error. MAE and RMSE are similar measures because they give estimates of the average error, but they do not provide information about the relative size of the average difference and the nature of differences comprising them. However, MAE and RMSE are among the best overall measures of model performance as they summarise the mean difference in the units of observed and predicted values. All of the measures assess the performance of spatial interpolation methods for individual primary variables. The magnitude of these measures depends on the unit/scale of the primary variable [24,25,45]:

$$ME = \frac{\sum_{i=1}^n Z^*(x_i) - Z(x_i)}{n} \quad (10)$$

$$MAE = \frac{\sum_{i=1}^n |Z^*(x_i) - Z(x_i)|}{n} \quad (11)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Z^*(x_i) - Z(x_i)}{Z(x_i)} \right| \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [(Z(x_i) - Z^*(x_i))]^2}{n}} \quad (13)$$

$$NSE = 1 - \frac{\sum_{i=1}^n [(Z(x_i) - Z^*(x_i))]^2}{\sum_{i=1}^n [(Z(x_i) - O)]^2} \tag{14}$$

$$PBAIS = 100 \cdot \frac{\sum_{i=1}^n Z(x_i) - Z^*(x_i)}{\sum_{i=1}^n (Z(x_i))} \tag{15}$$

where $z(x_i)$, $z^*(x_i)$, and O are the measured, interpolated, and mean of the observed values of water quality index of the i th well, respectively, while n is the sample size [22]. Nash–Sutcliffe efficiencies can range from $-\infty$ to 1. An efficiency of 1 ($E = 1$) corresponds to a perfect match of modeled discharge to the observed data [3]. An efficiency of 0 ($E = 0$) indicates that the model predictions are as accurate as the mean of the observed data, whereas an efficiency less than zero ($E < 0$) occurs when the observed mean is a better predictor than the model or, in other words, when the residual variance (described by the numerator in the expression above) is larger than the data variance (described by the denominator). The closer the model efficiency is to 1, the more accurate the model is [43].

4. Analysis of the interpolation methods and groundwater contamination

In order to analyze the effects of the parameters of the WDI model, on groundwater contamination, the weight indices of WDI is applied by using seven levels including 1–4 with a step of 0.5. For OK, the sample distribution should be normal, unless, a suitable change must be applied to the sampling data. The Kolmogorov-Smirnov test is used to test the normality of the distributions of water quality indices ($P < 0,05$) with XLStatsSoftware. The serial data in the WQI does not follow a normal distribution and hence, requires a log change. The model of optimal semivariogram has previously been fitted using Isatis 7 software, and corresponding parameters have been used for interpolation of Kriging on ArcGIS 10.2 software. The correlation is found between the Piezometric water level, which follows a normal distribution, and the WQI data series, in the OCK model. The GIS Maps produced by ArcGISsoftware using the following interpolation methods (IDW, OK, EBK, OCK) enables a comparison between different contamination levels. Sample data collected from wells and the WQI interpolated maps, are used to calculate the contaminated surface area. In this study, water quality indices (WQI) have been categories into five levels ($WQI < 50$, $50-100$, $100-200$, $200-300$ and $WQI > 300$), and classified as excellent quality, good, poor, very poor and not consumable, with respective indices values 1 to 5. The spatial analyst setting in the ArcGISsoftware is used, to convert the interpolated surface maps of WQI to raster pictures, next a subtraction is calculated between two assessed contaminated areas, in order to obtain several subtraction results [22].

In order to highlight this subtraction, several results are selected to evaluate the differences in the contaminated zone: where IDW4 is chosen to show the highest level of contamination in the zone, while OK and IDW1 is chosen to show the lowest level of contamination in the zone. IDW4

and IDW1 are used to show the impact of the weighting parameters on the assessment of contaminated zones [3].

5. Results

5.1. Interpolation values of WQI

Values given by descriptive statistics of water quality indices, in (Table 3) show a fluctuation of values between 37.6–348.36 with an average of 105.15. The higher values of water quality indices are higher in the North-East part of the plain. The main reasons being a higher density of population and the early settlement of industrial centers in this region. In additions, the trend is more or less the same in the different methods (Fig. 4). The Nugget coefficient ($Co/Co+C$) is used to classify spatial dependence under interpolation method of OK. Ratios of $< 25\%$, $25-75\%$ and $>75\%$ suggest auto-correlations, respectively high, middle, and low. The parameters of the best-fit exponential model for the data sets are given in (Table 4). The Nugget and Sill are very low. The Nugget value is 0.0385 and Sill value is 0.3063. The coefficient value of Nugget and Sill ($Co/Co+C$) of WQI is 0.1257. Nugget effect and Nugget-to-Sill ratio are used to classify the spatial dependence.

5.2. Accuracy of different interpolation methods

The precision of different interpolation methods is apparent in the indicators of crossed validation of WQI. For each interpolation method IDW, EBK OK, and OCK, the root means square error RMSE and the mean absolute error MAE provide a precise measure of interpolation, with the lowest values [22], while the ME measures the bias. In

Table 3
Statistic parameters of groundwater quality variables in Mitidja plain with the associated WQI

Groundwater	Mean (mg/l)	Min (mg/l)	Max (mg/l)	SD
Quality variable				
pH	7.84	7.39	8.20	0.19
Ca	140.65	51.95	227.00	49.62
Mg	43.49	17.98	92.96	19.69
Cl	161.23	33.00	442.50	111.60
SO ₄	175.56	15.00	428.75	104.99
NO ₃	38.78	0.00	84.85	21.57
CE*	1610.29	575.00	3300.00	700.98
TDS	1078.90	385.25	2211.00	469.66
TH °F	53.28	20.48	86.08	18.41
Fer	0.097	0.000	0.903	0.214
Manganese	0.080	0.000	1.965	0.326
Cuivre	0.019	0.000	0.052	0.014
Cadmium**	0.004	0.000	0.057	0.013
Plamb	0.004	0.000	0.016	0.005
WQI	105.15	37.60	348.36	77.84

*µS/cm; **µg/l

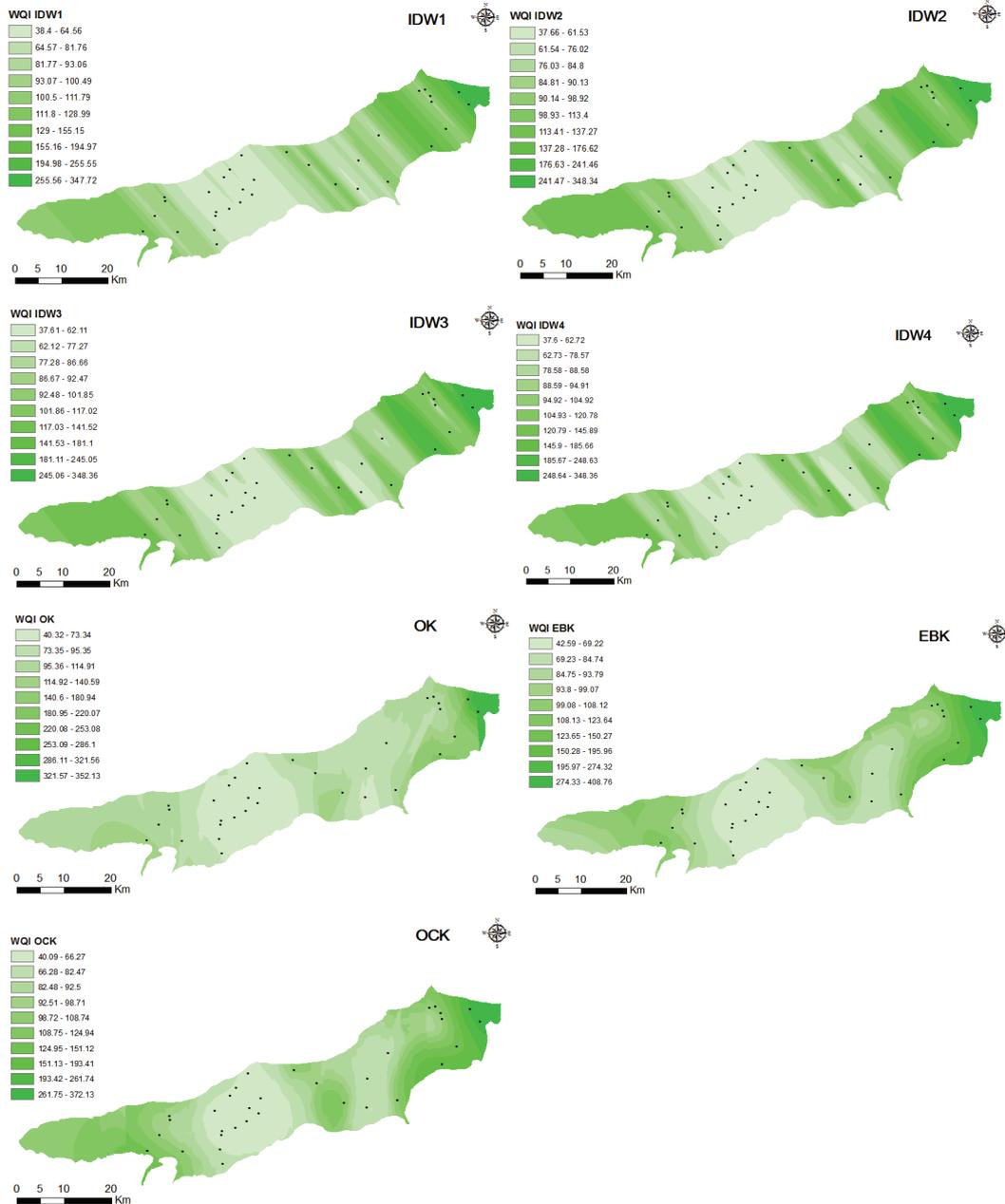


Fig. 4. Different interpolation methods and spatial prediction of WQI.

Table 4
Best fitted semivariogram model and parameters for the WQI dataset

Variable	WQI
Best-fit model	Gaussian
Nugget (C0)	0.0385
Sill (C + C0)	0.306
Range	14,745.75
Nugget/Sill	0.1257
R ²	0.9

Table 5, the RME varies between 35.63 and 42.12 where OCK has the best performance; next EBK shows the minimum difference between values. IDW4 show the maximum error value for both the weighted power and the root mean square error RMSE, proportionally, with the root mean square being the largest.

The bias represented by ME is almost equal to zero for all interpolation methods. OCK presents low values for MRE with a value of 0.277; other interpolation methods have the same value of 0.3. For Nash–Sutcliffe efficiency (NSE), OCK shows a robust performance with a value of 0.84, whereas IDW4 shows a weak performance off with a value of 0.776. OCK shows the best performance, whereas

Table 5
Prediction accuracy of the interpolation methods

Methods	ME	MAE	MRE	RMSE	NSE	PBAIS
OK	-4.8 E-05	30.47	0.321	40.57	0.792	4.58 E-05
OCK	6.1 E-04	26.98	0.277	35.63	0.840	-5.76 E-04
EBK	2.2 E-03	29.42	0.303	38.35	0.814	-2.12 E-03
IDW1	-1.6 E-05	27.82	0.296	40.37	0.794	1.52 E-05
IDW1.5	-7.0 E-05	28.10	0.295	41.49	0.782	6.66 E-05
IDW2	-1.6 E-05	28.08	0.295	41.60	0.781	1.52 E-05
IDW2.5	1.0 E-05	28.04	0.296	41.78	0.779	-9.56 E-06
IDW3	-1.8 E-05	27.97	0.296	41.94	0.778	1.74 E-05
IDW3.5	1.3 E-05	27.91	0.296	42.03	0.777	-1.22 E-05
IDW4	3.9 E-07	28.06	0.296	42.12	0.776	-3.74 E-07

EBK shows a weaker performance, with IDW showing the weakest performance. The PBAIS is approximately zero.

5.3. Average and variation coefficient of interpolation accuracy

All interpolation methods of WQI have similar average predictions. Average values are about 105.15 (Table 6). In contrast, variation coefficient varies between minimal values of 62.46 % for the OCK method to a maximum value of 76.05% for the IDW4 method. The simulated variation coefficient increases gradually as well as the weighted power increases for IDW interpolation method. The percentage value of variation coefficient shows that distribution is inconsistent which implies a high distribution around the average. Contributed by the size of the sampling site, the sample numbers collected, as well as the differences present in contamination areas [22].

5.4. Contaminated area and spatial distribution

In Table 7, percentages of polluted surfaces of each level vary from one interpolation method to another. In level 5 high contamination varies from 1.56% for EBK to 3.96% for OK. The contamination values of IDW is near to EBK which is a weaker variation between different interpolation methods. In contrary, the sample value (without interpolation) is higher (8.33%).

EBK and OCK show the lowest contamination area (level 5), with the lowest recorded percentage of (1.56% and 1.71% respectively), compared to other interpolation methods (Table 7). The variation of the polluted surface in IDW increases with the increase of the weighting level. The IDW1 registered 1.63% and the IDW4 recorded 1.95%.

The maps usually show the uncertainties in the projection of contamination surfaces [3,7,22]. Fig. 5 clearly shows the results of cell-by-cell subtraction between two raster images of two interpolation methods. The difference between the five levels of contamination, two by two, is represented by a color and a name “Low one level, Low two level, High one level, and High two-level” respectively from “Excellent water” to “water, unsuitable for drinking.” The subtraction between two raster images of the same level is expressed by

Table 6
The predicted mean and coefficient of variance using different methods

Methods	Predicted mean	Predicted coefficient of variation
OK	105.15	67.82
OCK	105.16	62.46
EBK	105.16	63.05
IDW1	105.15	68.38
IDW1.5	105.15	70.36
IDW2	105.15	71.89
IDW2.5	105.15	73.41
IDW3	105.15	74.52
IDW3.5	105.15	75.38
IDW4	105.15	76.05
Original value	105.15	74.02

Table 7
The water contamination area calculated by different methods

Methods	Groundwater contaminated area proportion (%)				
	Level 1	Level 2	Level 3	Level 4	Level 5
OK	2.06	41.41	47.28	5.28	3.96
OCK	6.52	50.83	38.94	2.00	1.71
EBK	1.63	40.68	52.30	3.83	1.56
IDW1	3.16	54.76	39.92	0.53	1.63
IDW1.5	5.41	54.30	37.69	0.87	1.72
IDW2	5.41	54.30	37.69	0.87	1.72
IDW2.5	6.49	52.00	38.96	0.76	1.79
IDW3	7.04	50.33	40.10	0.68	1.86
IDW3.5	7.33	48.90	41.24	0.62	1.91
IDW4	7.55	47.16	42.77	0.57	1.95
Simple ratio	16.67	47.22	27.78	0.00	8.33

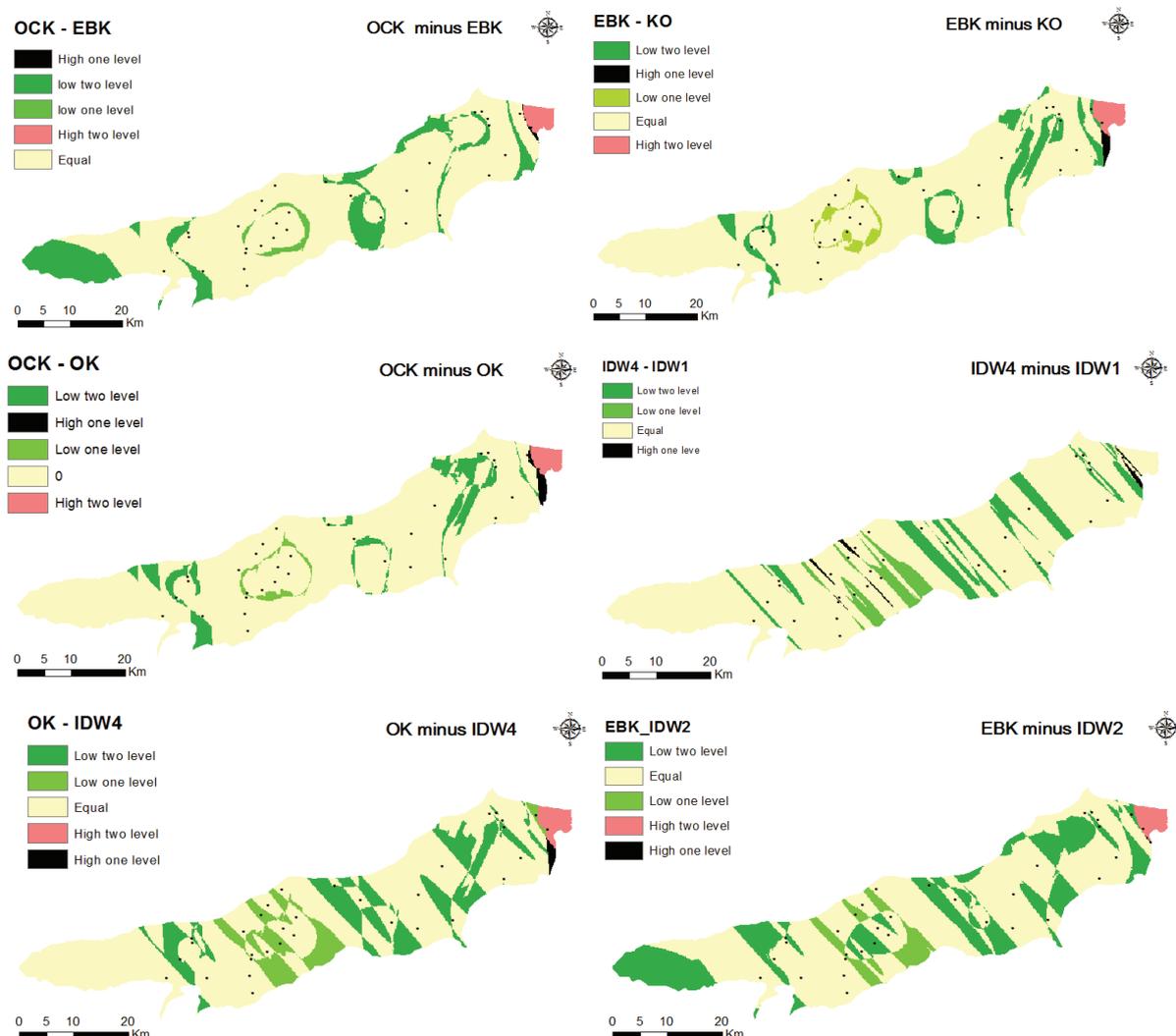


Fig. 5. The difference of WQI area estimated by each two interpolation techniques.

“Equal.” The results have more or less a similar level of contamination with a slight variation. The subtraction of EBK and OK are chosen to show the highest level of contamination in the area, and OCK and OK are selected to show the lowest level of contamination in the area. IDW4 and IDW1 are selected to assess the impact of weighting parameters on the contaminated area calculations. The spatial distribution of the polluted area estimates of OK – IDW4 and EBK – IDW2 are remarkably similar, which shows large uncertainties in the polluted areas. The results of IDW4 minus IDW1 show that IDW parameters expand the spatial scope in the center of Low one level and Low two-level values. The difference between OCK and EBK exists in the transition region from high to low WQI concentration.

6. Discussion

The WQI trend assessment by different interpolation methods is more or less the same with a slight change.

The maps produced by interpolation methods permit the contamination assessment of the Mitidja plain groundwater. This is realized by the influence of each method and its uncertainty. Liand Heap [25]; Xie et al., [22]; Kravchenko [44]; Liu et al. [4] explain the differences between methods by size and inconsistency of sampling data, the number of samples collected and the distance between sampling locations, as well as non-uniformity in contamination, by several types of contamination. Contamination is more present to the East (industrial zone related to metal processing) than in the Center and West of study plain. The precision of the methods depends on the calculation error from cross-validation which should be the lowest [3,22,45,46]. Biases of cross-validation in this study are extremely low for all interpolation methods, although a slight difference existed between RMSE, MRE, and NSE which gave more weight to interpolation by OCK followed by EBK (Table 5).

The calculation results of the contaminated surface area (Table 7) show that in EBK and OCK, level 1 and 2 contamination, estimate better than in other methods,

if we rely on the low errors to record when calculating MRE, RMSE and NSE and they represent the lowest Predicted coefficient of variation. Also, level 4 and 5 contaminations record the highest value evaluated by OK. These same levels increase respectively with IDW weighting. The same variation is noticed for the evolution of errors MRE, RMSE NSH and the predicted CV. In order to minimize errors of the global average, interpolation techniques allow smoothing of original data [22,45] with a purpose to preserve as far as a possible gradual variation of WQI values [3]. The variation of coefficient presented in (Table 6) shows that IDW presents the higher value than OCK and EBK. When the variation of coefficient decrease, errors of different interpolation methods also decrease. When variation coefficient increases, RMSE and MRE errors also increase [22].

The impact of weighting power of IDW varies proportionally with CV (Table 6) and (Table 7). Surface contamination also increases when the level of weighting increases. This result is concluded by Xie et al. [22] and Mirzaei et Sakizadeh, [3]. Subtraction between different interpolations methods shows that part of North-East, in Zone, contains the highest contamination level (Level 5). In addition, OCK is realized between water quality indices WQI and Piezometrical level of the same year, where probability results may arise in the interpolation method if a Piezometrical variation arises.

7. Conclusion

The consumption of chemically contaminated water, poses serious problem to public health. The national health and water authorities are always looking for the best method of assessing the contamination of groundwater and its spatial distribution for good groundwater management. This research is focused on the determination of the degree of groundwater contamination in the Mitidja Plain by selecting the most appropriate interpolation method to produce pollution maps. For that we analyzed the result of four spatial interpolation methods: EBK, IDW, OCK and OK compared to the water quality index (WQI). The interpolation methods tested had similar results for the prediction of Mitidja groundwater contamination. OCK RMSE, MRE, MAE and NSE estimation errors values are less than the EBK method. In addition, the IDW4 and the OK show the greatest error value, the higher the coefficient of variation, the greater is the error. Polluted zone assessment of levels 1 and 5 by a standard method, is greater than that realized by interpolation methods. OK, shows the largest polluted surface area of levels 4 and 5. The pollution is proportional to the weighting power of IDW. Uncertainty is shown in the transition zone between two levels. Level 5 contamination is present in the East of the plain, whereas level 4, the non-polluted zone, is located in the center of the plain. However, geostatistical interpolation methods (EBK, OK, OCK) require more analysis and perfection than *Deterministic interpolation* methods (IDW).

All the interpolation methods tested yielded almost similar predictions of the different levels of contamination. Regardless of the accuracy of the method, the sample size,

data quality and skills of the researchers will affect the uncertainty. The choice of an interpolation method for identifying various levels of contamination is not as simple as it seems.

Acknowledgment

The authors would like to thank the National Water Agency (ANRH) for its assistance and collaboration in providing the sample data need for our research article.

Symbols

C_i	— The concentration of each chemical parameter in each water sample in mg/L.
d_i	— The distance between point i and the prediction point.
n	— The parameters number. n is the sample.
q_i	— The quality scale for each parameter.
S_{li}	— The sub-index of i th parameter.
W_i	— The relative weight.
w_i	— The weight of each parameter.
WQI	— Water quality index
$Z(x)$	— The predicted value at an interpolated point.
$z(x_i), z^*(x_i), O$	— The measured, interpolated, and mean of the observed values of water quality index of the i th well, respectively.
$z(x_j)$ and $z^*(x_0)$	— The estimated value.
Z_i	— The amount at a known point.

References

- [1] J.-J. Lee, C.-S. Jang, S.-W. Wang, C.-W. Liu, Evaluation of potential health risk of arsenic-affected groundwater using indicator kriging and dose response model, *Sci. Total Environ.*, 384 (2007) 151–162.
- [2] S.M. Buchanan, J. Triantafyllis, Mapping water table depth using geophysical and environmental variables, *Ground Water*, 47 (2009) 80–96.
- [3] R. Mirzaei, M. Sakizadeh, Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran, *Environ. Sci. Pollut. Res.*, 23 (2016) 2758–2769.
- [4] C.-W. Liu, C.-S. Jang, C.-M. Liao, Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan), *Sci. Total Environ.*, 321 (2004) 173–188.
- [5] G. Gong, S. Mattevada, S.E. O'Bryant, Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas, *Environ. Res.*, 130 (2014) 59–69.
- [6] R. EgbuOtuIduma, T. Kingdom Simeon Abam, E. Daniel Uko, Geostatistical study of the spatial variability of groundwater parameters in Afikpo and Ohazara, Southeastern Nigeria, *J. Water Resour. Environ. Eng.*, 9(4) (2017) 72–85.
- [7] R. Liu, Y. Chen, C. Sun, P. Zhang, J. Wang, W. Yu, Z. Shen, Uncertainty analysis of total phosphorus spatial-temporal variation in the Yangtze River Estuary using different interpolation method, *Mar. Pollut. Bull.*, 86 (2014) 68–75.
- [8] H. Arslan, Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: the case of Bafra Plain, Turkey, *Agric. Water Manage.*, 113 (2012) 57–63.
- [9] B. Nas, A. Berkta, Groundwater quality mapping in urban groundwater using GIS, *Environ. Monit. Assess.*, 160 (2010) 215–227.

- [10] V. Kumar, Optimal contour mapping of groundwater levels using universal kriging—a case study, *Hydrol. Sci. J.*, 52 (2007) 1038–1050.
- [11] Z. Alizadeh, N. Mahjouri, A spatiotemporal Bayesian maximum entropy-based methodology for dealing with sparse data in revising groundwater quality monitoring networks: the Tehran region experience, *Environ. Earth Sci.*, 76 (2017) 436.
- [12] K. Khalili, Comparison of geostatistical methods for interpolation groundwater level (Case study: Lake Urmia Basin), *J. Appl. Environ. Biol. Sci.*, 4(1s) (2014) 15–23.
- [13] J. Joseph, H.O. Sharif, T. Sunil, H. Alamgir Application of validation data for assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents, *Environ. Pollut.*, 178 (2013) 411–418.
- [14] N. Theodossiou, P. Latinopoulos, Evaluation and optimization of groundwater observation networks using the Kriging methodology, *Environ. Model Softw.*, 21 (2006) 991–1000.
- [15] I. Triki, M. Zairi, H.B. Dhia, A geostatistical approach for groundwater head monitoring, network optimisation: case of the Sfax superficial aquifer (Tunisia), *Water Environ. J.*, 27 (2012) 362–372.
- [16] Y. Ran, X. Li, Y. Ge, X. Lu, Y. Lian, Optimal selection of groundwater-level monitoring sites in the Zhangye Basin, Northwest China, *J. Hydrol.*, 525 (2015) 209–215.
- [17] M. Keblouti, L. Ouerdachi, H. Boutaghane, Spatial interpolation of annual precipitation in Annaba-Algeria—comparison and evaluation of methods, *Energy Procedia.*, 18 (2012) 468–475.
- [18] E. BahramiJovein, S.M. Hosseini, Predicting saltwater intrusion into aquifers in vicinity of deserts using spatio-temporal kriging, *Environ. Monit. Assess.*, 189 (2017) 81.
- [19] C.C.F. Plouffe, C. Robertson, L. Chandrapala, Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: a case study of Sri Lanka, *Environ. Model Softw.*, 65 (2015) 57–71.
- [20] A. Izady, O. Abdalla, T. Ahmadi, M. Chen, An efficient methodology to design optimal groundwater level monitoring network in Al-Buraimi region, Oman, *Arab J. Geosci.*, (2017) 10–26.
- [21] A. Journel, P. Kyriakidis, S. Mao, Correcting the smoothing effect of estimators: a spectral postprocessor, *Math. Geol.*, 32 (2000) 787–813.
- [22] Y. Xie, T.-B. Chen, M. Lei, J. Yang, Q.-J. Guo, B. Song, X.-Y. Zhou, Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: accuracy and uncertainty analysis, *Chemosphere*, 82 (2011) 468–476.
- [23] P. Goovaerts, Estimation or simulation of soil properties? An optimization problem with conflicting criteria, *Geoderma*, 97 (2000) 165–186.
- [24] J. Li, A.D. Heap, A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors, *Ecol. Inform.*, 6(3–4) (2011) 228–241.
- [25] J. Li, A.D. Heap, spatial interpolation methods applied in the environmental science, *Environ. Model. Software*, 53 (2014) 173–189.
- [26] R.F. Siegel, *Environmental Geochemistry of Potentially Toxic Metals*. Springer, Berlin, 2002, 218 p.
- [27] G. Stamatis, Ground water quality of the Ag. Paraskevi Tempy valley karstic springs application of a tracing test for research of the micro-bial pollution (KatoOlympos/NE Thessaly), *Bull. Geol. Soc. Greece*, 43 (2010) 1868–1877.
- [28] G. Gnanachandrasamy, T. Ramkumar, S. Venkatramanan, S. Vasudevan, S.Y.M. Chung, Bagyaraj accessing groundwater quality in lower part of Nagapattinam district, Southern India: using hydrogeochemistry and GIS interpolation techniques, *Appl. Water Sci.*, 5 (2015) 39–55.
- [29] S.S. Dhindsa, P. Bheel, Y. Musturia, Hydrochemical study of ground water quality variation in Tonk District, Rajasthan, *Indian J. Environ. Ecolan.*, 8(1) (2004) 129–136.
- [30] V. Ramasubramanian, R. Jeyaprakash, D.A. Ruby Mallika, R. Ramasubbu, V. Mariappan, Analysis of physico-chemical characteristics of ground water quality and quality index in and around Sivakasi Town, *Indian J. Environ. Ecolan.*, 8(1) (2004) 171–176.
- [31] K.S. Murray, Hydrology and geochemistry of thermal waters in the Upper Napa Valley, California, *Ground Water*, 34 (1996) 1115–1124.
- [32] M. Rosen, S. Jones, Controls on the chemical composition of ground water from alluvial aquifers in the Wanaka and Wakatipu basins, CentralOtago, NewZealand, *Hydrogeol.*, 16 (1998) 264–281.
- [33] Minister des Ressources en Eau, l'agence de bassin Algérois – Houdna - Soummam Bassin de l'Algérois catnet N°1 (2002) 1–37. <http://www.abhahs.org/media/documents/carnet02.pdf>.
- [34] G. Mutin, Implantations industrielles et aménagements du territoire en Algérie. In: *Revue de géographie de Lyon*, 55 pp. 5–37; 1980
- [35] Y. Huh, G. Panteleyev, D. Babich, A. Zaitsev, J.M. Edmond, The fluvial geochemistry of the rivers of Eastern Siberia: II. Tributaries of the Lena, Omoloy, Yana, Indigirka, Kolyma, and Anadyr draining the collisional/accretionary zone of the Verkhoyansk and Cherskiy ranges, *Geochimica et Cosmochimica Acta*, 62(12) (1998) 2053–2075.
- [36] R.K. Horton, An index number system for rating water quality, *J. Water Contam. Control Fed.*, 37 (1965) 300–306.
- [37] M. Rupal, B. Tannushree, C. Sukalyan, Quality characterization of groundwater using water quality index in Surat City, Gujarat, India, *Int. Res. J. Environ. Sci.*, 1(4) (2012) 14–23.
- [38] A.K. Tiwari, P.K. Singh, M.K. Mahato, GIS-Based Evaluation of water quality index of groundwater resources in West Bokaro coalfield, India, *Current World Environ.*, 9(3) (2014) 843–850.
- [39] M. Sakizadeh, E. Ahmadpour, Geological impacts on groundwater pollution: a case study in Khuzestan Province, *Environ. Earth Sci.*, (2016) 75–88.
- [40] V.P. Samsonova, N. Blagoveshchenskii Yu, L. Meshalkina Yu, Use of empirical Bayesian Kriging for revealing heterogeneities in the distribution of organic carbon on agricultural lands, *Eurasian Soil Sci.*, 50(3) (2017) 305–311.
- [41] K. Krivoruchko (a) Empirical Bayesian Kriging. ArcUser Fall (2012) 6–10. Also available online at <http://www.esri.com/news/arcuser/1012/empirical-bayesian-kriging.html>.
- [42] T. Heisel, K.A. Ersboll, C. Andreasen, Weed mapping with Co-Kriging using soil properties, *Precision Agri.*, 1 (1999) 39–52.
- [43] P.D. Wagner, P. Fiener, F. Wilken, S. Kumar, K. Schneider, Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions, *J. Hydrol.*, 464–465 (2012) 388–400.
- [44] a) A.N. Kravchenko, Influence of spatial structure on accuracy of interpolation methods, *Soil Sci. Soc. Am. J.*, 67 (2003) 1564–1571. b) K. Krivoruchko, Modeling Contamination Using Empirical Bayesian Kriging. ArcUser Fall 2012 <http://www.esri.com/news/arcuser/1012/modeling-contamination-using-empirical-bayesian-kriging.html>.
- [45] O. Falivene, L. Cabrera, R. Tolosana-Delgado, A. Sáez, Interpolation algorithm ranking using cross-validation and the role of smoothing effect. A coal zone example, *Comput. Geosci.*, 36 (2010) 512–519.
- [46] K. Johnston, J.M. Ver Hoef, K. Krivoruchko, N. Lucas, Using ArcGIS Geostatistical Analyst (ESRI Userbook). (2001) and (2003) http://dusk2.geo.orst.edu/gis/geostat_analyst.pdf; http://downloads2.esri.com/support/documentation/ao_/Using_ArcGIS_Geostatistical_Analyst.pdf.