

Comparison of Spatial Interpolation Methods for Mapping Seasonal Groundwater Levels



Akash Singh Raghuvanshi and H. L. Tiwari

Abstract Proper knowledge of spatio-temporal variation of groundwater level are very important for efficient planning and management of groundwater level. In sparsely monitored basins, levels of groundwater are generally monitored at random points which may be far away from each other. Interpolation of groundwater level over a region in an accurate manner is a pre-requisite for modeling as well as management of water resources that can be achieved by adopting accurate and reliable techniques of interpolation of scattered data over a region. In the present study, two deterministic interpolation methods, i.e., inverse distance weighted (IDW), radial basis functions (RBFs), and three geostatistical interpolation methods, i.e., ordinary kriging (OK), simple kriging (SK), and ordinary cokriging (OCK) are evaluated to predict the spatial and temporal variation of groundwater levels. Cross-validation procedure is adopted to evaluate the prediction performance of adopted interpolation methods. Groundwater level data for the year 2019 from 31 sparsely located monitored observation wells over Sagar district of Madhya Pradesh, India are used to evaluate the performance of interpolation methods with respect to different statistical indicators of cross-validation. Interpolation maps of estimate groundwater level are produced using all five spatial interpolation techniques. Results of the analysis are presented and discussed.

Disclaimer: The presentation of material and details in maps used in this chapter does not imply the expression of any opinion whatsoever on the part of the Publisher or Author concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its borders. The depiction and use of boundaries, geographic names and related data shown on maps and included in lists, tables, documents, and databases in this chapter are not warranted to be error free nor do they necessarily imply official endorsement or acceptance by the Publisher or Author.

A. S. Raghuvanshi (✉) · H. L. Tiwari
Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee 247667,
Uttarakhand, India
e-mail: akash_sr@hy.iitr.ac.in

A. S. Raghuvanshi
Department of Civil Engineering, Maulana Azad National Institute of Technology, Bhopal
462003,, Madhya Pradesh, India

Keywords Groundwater level · Spatial interpolation · Geostatistics · Cross-validation · ArcGIS 10.8

1 Introduction

Groundwater is one of the most important water sources. Groundwater management is important for their sustainable development. Therefore, we need appropriate information about the spatio-temporal behavior of the water table in a region. However, monitoring groundwater levels is inherently costly and time consuming, especially during installation stages that require the drilling of wells and piezometers. As a result, the number of surveillance sites available in a particular region is relatively small and often does not reflect the actual range of variation that may exist. Therefore, accurate groundwater level spatial interpolation at unsampled sites is required for better groundwater management.

Among various interpolation methods, there is no clear optimal method, thus results need to be compared depending on specific situation [10]. This study presents the use of GIS tools to generate the groundwater level surface for a sparsely monitored region from groundwater levels monitored at random locations. Geostatistical analyst tool in ArcGIS 10.8 is used to explore the spatial variability in groundwater levels for the Sagar district region located in Madhya Pradesh, India.

2 Study Area and Data Source

2.1 Sagar District

The Sagar district is located in the north central part of Madhya Pradesh, India, and occupies an area of 10,252 km². The district extends between the latitude of 23° 10' and 24° 27'N, longitude of 78° 04' and 79° 21'E. Figure 1 shows the index map for the Sagar district.

2.2 Data Used

The pre (April–June) and post-monsoon (Oct–Dec) seasonal groundwater level data of 31 central ground water board (CGWB) monitoring wells for 2019 are used. The location and groundwater level of observation wells are collected directly from the India Water Resources Information System (WRIS). Traditionally, geostatistical studies are performed on at least 100 samples [11]. However, the sample size is small in this study.

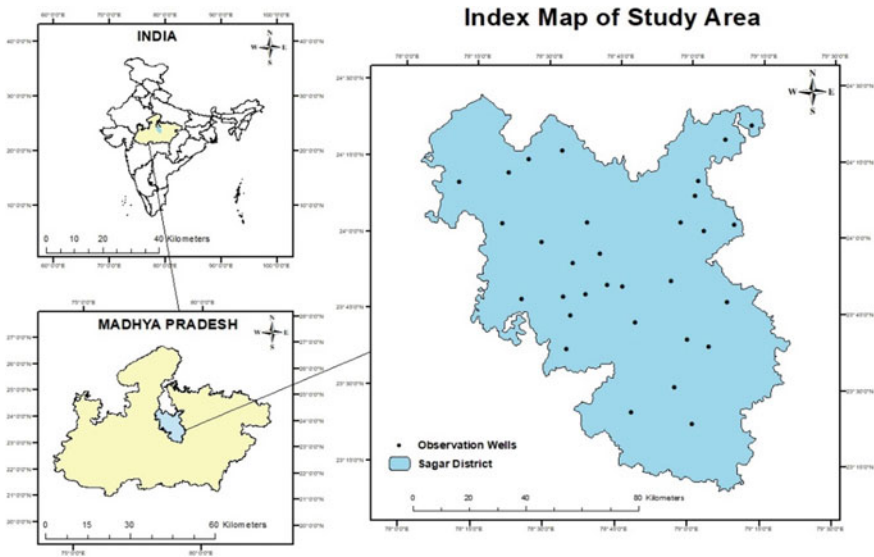


Fig. 1 Index map of study area

3 Materials and Method

3.1 Exploratory Analysis

The datasets are initially visualized in order to identify incorrect coordinate information and illogical data points. The screened datasets are then subjected to exploratory information evaluation to identify the outliers which can be unfavorable to spatial prediction. The variogram, in particular, may be very touchy to outliers due to the fact it's far primarily based totally at the squared differences among information [8]. Description of the information values is done through fundamental precis of statistics, inclusive of means, medians, variances, and skewness.

The geostatistical method mainly kriging is taken into consideration the high-quality unbiased linear prediction (BLUP) if the information meets the situations of normality, variance uniformity, and stationarity [7]. However, spatial information, specifically weather information, violates those situations. High asymmetries and outliers have undesired consequences on variogram shape and kriging estimates [6]. For spatio-temporal data that follow a Gaussian distribution, the effects of extremes are reduced, and more stable variograms are obtained, making it easier to model spatial variability [5]. Data transformations is required prior to kriging to standardize data distribution, eliminate outliers, and improvise data stationarity [4]. In this study, the normality of groundwater level data is visualized using tools such as histograms and boxplots, and the mean and median, symmetry (skewness) and

flattening (kurtosis) coefficients. It is checked numerically by comparing it with the normal distribution and also according to the formal statistical Shapiro–Wilk test.

3.2 Interpolation Methods

Inverse distance weighting directly works on the assumption that the point closest to the predicted position is weighted more heavily, and the weight is reduced as a function of distance, hence the name inverse distance weighting. The exact formula for this interpolator is

$$z(s_0) = \frac{\sum_{i=1}^n \frac{s_i}{d_i^p}}{\sum_{i=1}^n \frac{1}{d_i^p}} \quad (1)$$

where $z(s_0)$ is predicted value of unsampled point, n is total no. of sampling points, d_i is separation distance between unsampled point and i th sampled point, and p denotes the weighing power.

Radial basis functions (RBFs) are an exact set of interpolation methods. That is, the surface must pass through each measurement sample value. RBF is used to generate smooth surfaces from a large number of data points. There are five different functions: thin-plate spline (TPS), spline with tension (SPT) [2], fully regularized spline (CRS) [2], multiquadric function (MQ), and inverse multiquadric function (IMQ) [12]. Each basis function has a different form and has a different interpolation plane.

Kriging belongs to a family of generalized least squares regression methods in geostatistics which uses observations sampled in a particular search environment as a linear combination to estimate values in unsampled locations [5, 7].

$$\widehat{Z}(s_o) = \sum_{\alpha} \omega_{\alpha} Z(s_{\alpha}) \quad (2)$$

$\widehat{Z}(s_o)$ is the estimated variable of interest (groundwater level) at the unsampled location s_o , and $Z(s_{\alpha})$ is the observed values at the sampled locations in the vicinity of s_o .

OK is most commonly used form of kriging, in which mean is considered unknown and fluctuates locally, which makes it possible to maintain stationarity in the local neighborhood.

Simple kriging is mathematically the simplest but the least common. It is assumed that the expected value (mean) of the random field is known and depends on the covariance function. SK assumes second-order stationary that is constant mean, variance, and covariance across the domain or the region of interest [11].

The OCK method is a modification of the OK method. The main advantage of OCK is that it uses multiple variables in the estimation process. The OCK method

is used to improve the predictability of the primary variable by using the auxiliary variable, assuming that both primary and auxiliary variables are in good correlation [7]. This method is especially suitable when the main attributes of interest are sparse, but the relevant secondary information is abundant.

3.3 Variograms (Semivariogram)

The existence of spatial structures (spatial autocorrelation) in which nearby observations are more similar than distant observations is a pre-requisite for the application of geostatistics [5]. Experimental variograms measure the average dissimilarity between unsampled values and nearby data values [4] and can therefore represent autocorrelation at various distances. The kriging method requires a model of functions that characterize spatial variability, variograms, and key characteristic parameters such as nugget effects, thresholds, and ranges [5]. The experimental variogram is calculated based on the following formula:

$$\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{\alpha=1}^{N(d)} (Z(s_{\alpha} + d) - Z(s_{\alpha}))^2 \quad (3)$$

with $Z(s_{\alpha})$ and $Z(s_{\alpha} + d)$ being the values observed at the locations u_{α} and $u_{\alpha} + d$ separated by the distance d and $N(d)$ being the number of such pairs. If the value at $Z(s_{\alpha})$ and $Z(s_{\alpha} + d)$ is autocorrelated, the result of Eq. (3) will be small, relative to an uncorrelated pair of points. From an experimental variogram analysis, the appropriate model (Gaussian, spherical, etc.) is usually fitted by the weighted least squares method, and the parameters (range, threshold, and nugget) are used in kriging. Exponential, Gaussian, and spherical are the most commonly used (theoretical) variogram models in hydrological kriging applications [1] and are also used to model experimental variograms.

3.4 Cross-Validation

The performance of the various interpolation methods (IDW, RBF, SK, OK, and OCK) is evaluated through a cross-validation process. Cross-validation is a validation technique that removes observations one by one from the dataset and re-estimates from the remaining sampled data using the selected model. If the sample size of the data is very small, such as when there are only 31 observations, the method comparison is done by cross-validation [7]. This is a common way to verify the accuracy of the interpolation method [1].

The overall performance of the interpolation methods for groundwater level estimation is conducted using correlation and error-based measures. The correlation

includes the coefficient of determination (R^2), Nash–Sutcliffe efficiency coefficient (E), and Willmott agreement index (d), whereas the error measures include the mean relative error (MRE), the root mean square error (RMSE), and the mean error (ME).

4 Results and Discussion

4.1 Exploratory Data Analysis

Seasonal groundwater data from 31 observation wells are examined to understand the pattern in the data. Histogram (Fig. 2) and standard statistics (Table 1) are used to describe the data.

The histogram shows symmetry indicating normality of distribution. In fact, seasonal groundwater elevation data have mean and median close enough and skewness value close to zero, thus suggesting this distribution as symmetrical distribution. Finally, the Shapiro–Wilk test confirms the normality of the original data ($p = 0.83 > 0.05$, $p = 0.71 > 0.05$). Since original data follow a normal distribution hence, it is decided to work on original data without transformation.

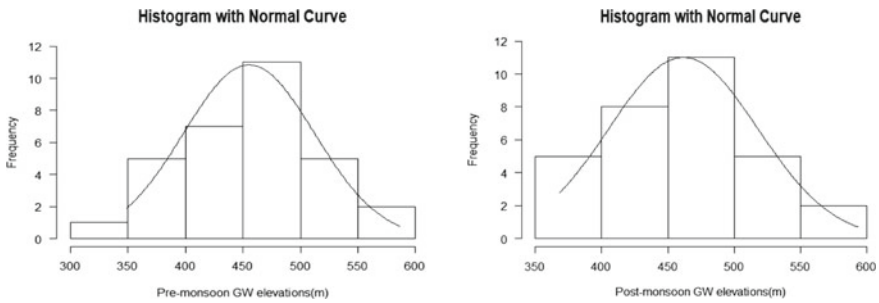


Fig. 2 Histogram of seasonal groundwater levels (amsl). Curve represents the fitting of a normal distribution

Table 1 Descriptive statistics for seasonal groundwater elevation level (m) data

Season	Min	Max	Mean	SE	Median	Q1	Q3	SD	CV	Skew	Kurt	SW
Pre-monsoon	348.9	586.1	455.3	10.2	460.1	416.1	490.9	57.1	12.53	0.10	– 0.64	0.83
Post-monsoon	368.4	592.9	461.7	10.1	464.0	420.1	496.6	56.2	12.17	0.13	– 0.61	0.71

Min. minimum, Max. maximum, SE. standard error, Q1 1st quartile, Q3 3rd quartile, SD standard deviation, Skew. skewness, Kurt. kurtosis, CV coefficient of variation, SW probability corresponding to the Shapiro–Wilk test of normality

4.2 Spatial Analysis of Groundwater Level Data

The spatial variation of seasonal groundwater elevation is considered to be isotropic ignoring the separation direction because the size of the sample (31) is limited and would not possibly detect anisotropy [7].

Table 3 and Figs. 3 and 4 show the variogram model and its parameters tuned for seasonal groundwater levels. The nugget effect to nugget to threshold ratio is used to classify the spatial dependence of the variable [3]. Variables have strong spatial dependence when the nugget-threshold ratio is less than 0.25, and moderate spatial dependence when the ratio is 0.25 to 0.75 [8]. Otherwise, the variables are less spatially dependent. Therefore, in our case, the groundwater level was strongly spatially correlated (Table 2).

Elevation, as auxiliary information, has decreased semi-variances. It may be visible that the sill is higher for OK and SK (4507.4 m² and 4,295.7 m²) than for OCK (3713.8 m² and 3577 m²). This is predicted due to the fact the covariate, elevation which become taken into consideration for OCK, however, now no longer for OK and SK variogram, in part explains the variability of the groundwater level data.

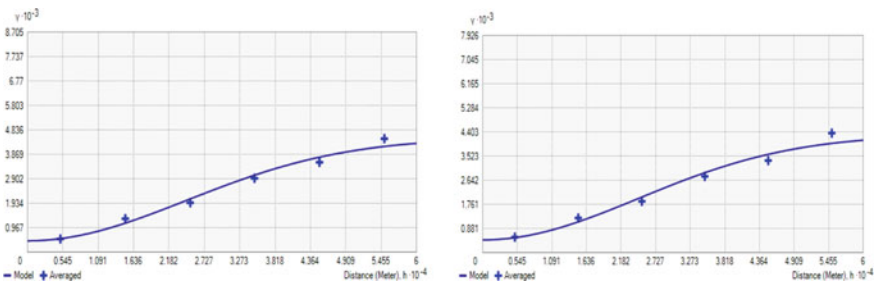


Fig. 3 Experimental (points) and fitted theoretical (curve) variograms of pre-monsoon (left) and post-monsoon (right) groundwater levels for OK and SK

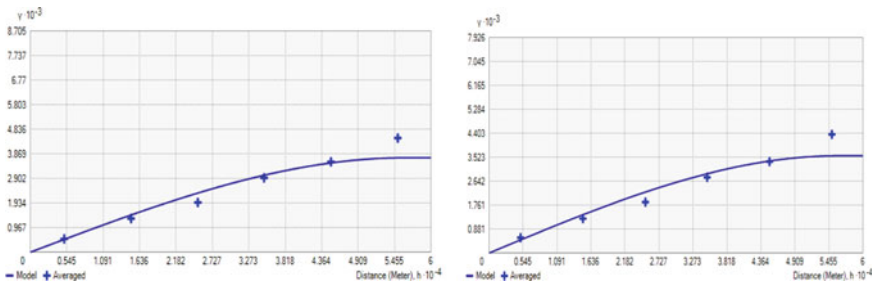


Fig. 4 Experimental (points) and fitted theoretical (curve) variograms of pre-monsoon (left) and post-monsoon (right) groundwater levels for OCK

Table 2 Summary of semivariogram parameters of best-fitted theoretical model to predict seasonal groundwater level (amsl)

Interpolation	Season	Best-fit model	Nugget (m)	Sill (m)	Range (km)	Nugget/Sill
OK & SK	Pre-monsoon	Gaussian	449.5	4507.4	60	0.1
	Post-monsoon	Gaussian	465.0	4,295.7	60	0.11
OCK	Pre-monsoon	Spherical	0	3713.8	56	0
	Post-monsoon	Spherical	0	3577	56	0

Table 3 Summary of descriptive statistics for observed and predicted pre-monsoon groundwater level

Pre-monsoon	Mean	Max	Min	Median	C5	C25	C75	C95	SD	CV
Observed	455.31	586.18	348.93	460.15	369.9	416.15	490.9	539.44	57.09	12.53
IDW	459.20	524.98	389.37	454.86	408.47	426.34	493.52	505.80	36.82	8.02
RBF	455.96	531.87	355.74	463.84	383.73	419.81	491.70	524.13	46.29	10.15
OK	456.16	525.41	367.55	452.66	383.88	424.14	493.96	520.94	43.97	9.64
SK	456.34	526.9	373.19	452.65	384.13	426.06	494.13	521.84	43.90	9.62
OCK	456.10	573.65	369.1	464.45	382.38	414.46	491.36	528.43	51.58	11.31

For IDW method, an optimal power value (p) as well as the number of the closest neighbors to include are determined, whereas for RBF method, the choice of radial basis function, their kernel parameter, and number of the closest neighbors to include are determined by minimizing the root mean square error (RMSE) statistics obtained from a cross-validation procedure. In this study, for IDW method, the optimizing parameter of the weight function (p) is taken as 2.0, whereas for RBF method, multiquadric radial basis function is found to be optimal among all the functions.

4.3 Groundwater Mapping

Tables 3 and 4 show various descriptive statistical parameters for the measured seasons (before and after the monsoon) and the parameters predicted by two deterministic and three geostatistical interpolation methods.

One of the hallmarks of geostatistical methods is smoothing, as predicted values are less variable than measured values. In other words, the minimum expected value is greater than the measured value, and the maximum expected value is less than the measured value [9]. This smoothing phenomenon is least for OCK while it is the most accentuated for SK which has 373.19, 376.34 m as minimal and 526.9, 533.14 m as maximal values compared to 348.93, 368.43 m as minimal and 586.18, 592.98 m as maximal for the measured groundwater values during pre- and post-monsoon periods, respectively.

Table 4 Summary of descriptive statistic for observed and predicted post-monsoon groundwater level

Post-monsoon	Mean	Max	Min	Median	C5	C25	C75	C95	SD	CV
Observed	461.68	592.98	368.43	464.05	373.57	420.1	496.65	547.28	56.18	12.17
IDW	465.25	531.62	390.41	460.35	417.72	431.36	500.74	513.31	37.29	8.01
RBF	461.87	538.51	358.2	468.28	394.93	426.49	497.84	529.61	45.84	9.92
OK	462.21	525.41	367.55	454.70	395.24	429.62	499.88	527.09	43.78	9.47
SK	462.51	533.14	376.34	454.70	395.56	431.98	500.88	528.03	43.60	9.42
OCK	462.14	580.27	377.8	469.88	384.66	420.20	497.92	535.21	51.61	11.16

Min. Minimum, Max. maximum, C5 5th centile, C25 25th centile or 1st quartile, C75 75th centile or 3rd quartile, C95 95th centile, SD standard deviation, CV coefficient of variation, OK ordinary kriging, SK simple kriging, OCK ordinary cokriging

This phenomenon is confirmed by the standard deviation, especially the decrease in the estimated variance for the measured data variance of 43, 39.7% of SK; 40.6, 39.2% for OK; 18.4, 15.6% for OCK during pre- and post-monsoon periods, respectively, and also by the coefficient of variation which is minimal for SK (9.62, 9.42%) followed by OK (9.64, 9.47) and OCK (11.31, 11.16%) compared to measured values (12.53, 12.17%) for pre- and post-monsoon periods.

Smoothing phenomenon among deterministic methods is least for RBF and most accentuated for IDW which has 389.37, 390.41 m as minimal and 524.98, 531.62 m as maximal values compared to 348.93, 368.43 m as minimal and 586.18, 592.98 m as maximal for the measured groundwater values during pre- and post-monsoon periods, respectively.

In geostatistical methods, estimates of SK are higher than those of OCK and OK with SK, OK, and OCK mean values of 456.34, 462.51 m; 456.16, 462.21 m; 456.10, 462.14 m for pre- and post-monsoon periods, respectively. Similar results can be seen in minimal values and for different percentiles. Similar analysis is carried out for deterministic interpolation methods, and overall estimates of IDW are found to be higher among all five interpolation methods used in this study.

Figures 5 and 6 represent the groundwater maps obtained from five methods of spatial interpolation (IDW, RBF, SK, OK, and OCK) for pre- and post-monsoon periods for the year 2019.

4.4 Performance Evaluation Study

To deepen the comparative study of spatial interpolation, the performance indicators for cross-validation are shown in Figs. 7 and 5. The boxplot of groundwater level prediction error (Fig. 7) shows that interpolation generally corrects the predicted groundwater level of 31 observation wells. The perfect match between the predicted and measured values is represented by 0 in the Fig. 7. Comparing the five methods, the residual (error) between the measured and predicted groundwater levels of OCK

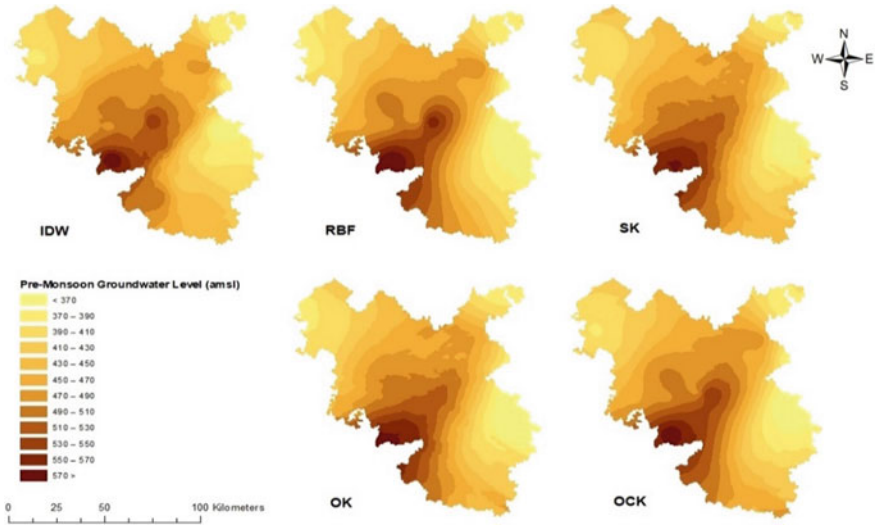


Fig. 5 Maps of pre-monsoon groundwater level (m) amsl estimated by IDW, RBF, SK, OK, and OCK

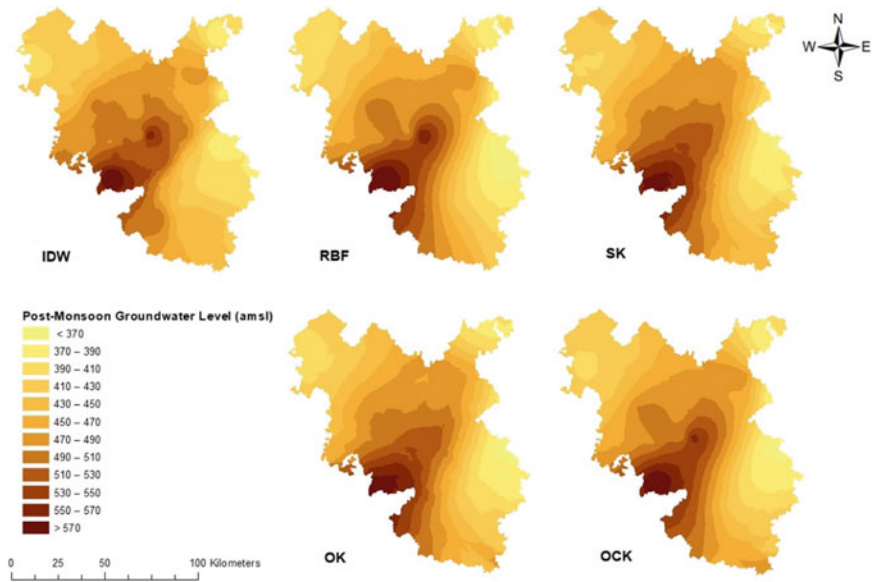


Fig. 6 Maps of post-monsoon groundwater level (m) amsl estimated by IDW, RBF, SK, OK, and OCK

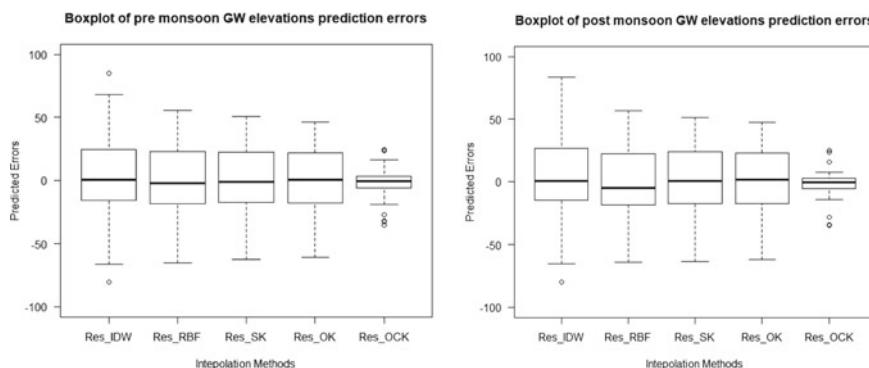


Fig. 7 Boxplots of seasonal groundwater level prediction errors using IDW, RBF, SK, OK, OCK

is significantly reduced. This shows that OCK is the best interpolator, though there exists small underestimation and overestimation.

Because aim of this study was to compare various methods, first the impact of various interpolation method on accuracy was considered. The 3 kriging methods performed better than the deterministic approach to estimate groundwater levels for each season. Performance measures of interpolation methods are summarized in Table 5. High values of coefficients of determination, Nash–Sutcliffe efficiency, and Willmott agreement index suggested an amazing fit among measured and predicted water level depth. Of the five interpolation methods, OCK had the best overall performance, with OK, SK, and RBF significantly superior to IDW. Not only the performance indicators of the model, but also the errors confirmed the above facts. The low RMSE and ME of all interpolation methods showed applicability to the prediction of groundwater level, and the superiority of OCK over all other methods was fully demonstrated by its minimum error value.

The second approach to assess the accuracy of a method is done by adopting regression coefficients (intercept and slope). The best model performance is represented by small intercept and large gradient. Among the five interpolation methods, the OCK, method which considered elevation as an auxiliary variable for predicting groundwater levels, showed best results.

5 Conclusions

The following conclusions are derived from the foregoing study:

- Introduction of elevation information improved the performance of covariate kriging method, OCK in particular, in sparsely monitored region.

Table 5 Performance evaluation of interpolation methods to predict groundwater levels

Interpolation	Season	Efficiency			Error			Intercept	Slope
		R ²	E	d	RMSE	ME	MRE		
IDW	Pre-monsoon	0.62	0.60	0.83	35.51	3.88	0.14	226.73	0.51
	Post-monsoon	0.63	0.61	0.84	34.49	3.56	0.15	221.66	0.52
RBF	Pre-monsoon	0.72	0.72	0.90	29.44	0.65	0.12	141.14	0.69
	Post-monsoon	0.72	0.72	0.90	29.20	0.18	0.13	141.82	0.69
SK	Pre-monsoon	0.75	0.74	0.91	28.58	1.02	0.12	152.95	0.66
	Post-monsoon	0.75	0.74	0.91	28.01	0.83	0.12	151.89	0.67
OK	Pre-monsoon	0.77	0.76	0.92	27.56	0.85	0.11	148.23	0.67
	Post-monsoon	0.76	0.76	0.92	27.17	0.52	0.12	146.98	0.68
OCK	Pre-monsoon	0.98	0.98	0.99	5.82	0.25	0.02	49.587	0.89
	Post-monsoon	0.99	0.99	0.99	4.79	- 0.01	0.02	42.536	0.90

- Smoothing phenomenon in geostatistical method is least for OCK while it is the most accentuated for SK whereas smoothing phenomenon among deterministic methods is least for RBF and most accentuated for IDW.
- In geostatistical methods, estimates of SK were higher than OCK and OK, whereas overall estimates of IDW are found to be higher among all five interpolation methods.
- Geostatistical methods performed better than the deterministic methods for predicting groundwater levels.

References

1. Adhikary PP, Dash CJ (2017) Comparison of deterministic and stochastic methods to predict spatial variation of groundwater depth. *Appl Water Sci* 7:339–348
2. Arslan H (2014) Estimation of spatial distribution of groundwater level and risky areas of seawater intrusion on the coastal region in Carsamba Plain, Turkey, using different interpolation methods. *Environ Monit Assess* 186:5123–5134
3. Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field-scale variability of soil properties in Central Iowa Soils. *Soil Sci Soc Am J* 58:1501–1511
4. Deutsch CV, Journel AG (1998) *GSLIB: Geostatistical Software Library and user's guide*, 2nd edn. Oxford University Press, New York
5. Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, New York
6. Gringarten E, Deutsch CV (2001) Teacher's aide: variogram interpretation and modeling. *Math Geol* 33:507–534
7. Isaaks EH, Srivastava RM (1989) *An introduction to applied geostatistics*. Oxford University, New York

8. Liu D, Wang Z, Zhang B, Song K, Li X, Li J, Li F, Duan H (2006) Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. *Agric, Ecosyst Environ* 113(1–4):73–81
9. Rata M, Douaoui A, Larid M (2020) Comparison of geostatistical interpolation methods to map annual rainfall in the Chélif watershed, Algeria. *Theoret Appl Climatol* 141:1009–1024
10. Sun Y, Kang S, Li F, Zhang L (2009) Comparison of interpolation methods for depth to groundwater and its temporal and spatial variation in the Minqin oasis of Northwest China. *Environ Model Softw* 24:1163–1170
11. Webster R, Oliver MA (2007) *Geostatistics for environmental scientists*, 2nd edn. Wiley, Chichester
12. Xie Y, Chen T, Lei M, Yang J (2011) Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: accuracy and uncertainty analysis. *Chemosphere* 82:468–476