A new trend function-based regression kriging for spatial modeling of groundwater hydraulic heads under the sparse distribution of measurement sites

S. Mohanasundaram[1] · Parmeshwar Udmale[1] · Sangam Shrestha[1] · Triambak Baghel[1] · Smit Chetan Doshi[1] · Balaji Narasimhan[2] · G. Suresh Kumar[3]

Abstract

Discrete groundwater level datasets are interpolated often using kriging group of models to produce a spatially continuous groundwater level map. There is always some level of uncertainty associated with different interpolation methods. Therefore, we developed a new trend function with the mean groundwater level as a drift variable in the regression kriging approach to predict the groundwater levels at the unvisited locations. Groundwater level data for 29 observation wells in Adyar River Basin were used to assess the performance of the developed regression kriging models. The cross-validation results shows that the proposed regression kriging method in the spatial domain outperforms other physical and kriging-based methods with R^2 values of 0.96 and 0.98 during pre-monsoon and post-monsoon seasons, respectively.

Keywords Regression kriging · Digital elevation model · Mean groundwater level · Geostatistics · Trend function · Prediction error variance

Introduction

The prediction of groundwater hydraulic heads or groundwater levels is vital for the sustainable management of groundwater resources. In the temporal domain, the groundwater levels are predicted broadly by the deterministic physical models or stochastic time series models. Stochastic time series modeling methods are widely applied for predicting the groundwater levels as they require limited data, unlike physical-based models that require intensive data and parameterization process. The stochastic time series modeling approach can be broadly grouped under univariate and multivariate time series modeling methods. There have been

many applications of univariate and multivariate time series modeling approaches for predicting the groundwater levels (Bierkens et al. 1999, 2001; Tankersley et al. 1993; Shirmohammadi et al. 2013; Mohanasundaram et al. 2017, 2019). In spatial domain, the kriging approach is widely used in environmental studies to interpolate the point-based discrete variables such as rainfall, soil moisture, groundwater levels, surface elevation, air temperatures, soil and water quality parameters at the unknown locations (Theodossiou and Latinopoulos 2006; Dash et al. 2010; Wagner et al. 2012; Arun 2013; Wu and Li 2013; Gong et al. 2014; Vereecken et al. 2014; Plouffe et al. 2015; Guekie simo et al. 2016; Landrum et al. 2016; Martínez-Murillo et al. 2017; Rizo-Decelis et al. 2017; Shtiliyanova et al. 2017; Theodoridou et al. 2017; Varouchakis et al. 2018; Yin et al. 2019; Amini et al. 2019).

The groundwater levels mostly follow a non-stationary process and exhibit a strong trend pattern in the spatial domain due to varying hydrogeological, hydrological and climatic conditions (Varouchakis and Hristopulos 2013). Therefore, the spatial trend component in the groundwater levels must be accounted for a suitable trend function while interpolating the groundwater levels at unknown locations. The spatial trend component of the groundwater levels can

✉ S. Mohanasundaram
mohanasundaram@ait.asia

[1] Water Engineering and Management, Department of Civil and Infrastructure Engineering, School of Engineering and Technology, Asian Institute of Technology, Pathum Thani 12120, Thailand

[2] Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600036, India

[3] Department of Ocean Engineering, Indian Institute of Technology Madras, Chennai 600036, India

be modeled with various mathematical functions by different kriging group of models. For example, the spatial trend component can be modeled with a known stationary constant mean by simple kriging (SK) approach (Sun et al. 2009). Alternatively, if the mean of the variable is unknown, the data can be modeled by ordinary kriging (OK) approach (Ahmadi and Sedghamiz 2008; Varouchakis et al. 2012; Daya and Bejari 2015; Adhikary and Dash 2017). The universal kriging (UK) approach is adopted if a regional gradient present in the modeled datasets (Kumar 2007). Sometimes, the spatial trend in the groundwater levels can be explicitly modeled in terms of auxiliary variables such as digital elevation model (DEM), distance to stream network (DS), and other explanatory variables. This group of kriging is called regression kriging (RK) where groundwater levels at the observation locations are first detrended with appropriate deterministic trend functions and the remaining random residuals are interpolated using OK method (Rivest et al. 2008; Varouchakis and Hristopulos 2013; Zhu et al. 2013; Mhamad 2019). The overall accuracy in the groundwater water level predictions is relatively better for RK methods as it separately filters the spatial trend component from the groundwater levels which, in turn, reducing the variance of the residual component when compared to SK and OK methods (Varouchakis and Hristopulos 2013). Many studies have correlated the auxiliary variables such as spatial coordinates, topographic index, DS, and DEM with groundwater levels in the trend functions of RK models (Desbarats et al. 2002; Nikroo et al. 2010; Chung and Rogers 2012; Varouchakis and Hristopulos 2013; Zhu et al. 2013). However, the overall uncertainty which is arising from RK model interpolations can be effectively minimized by minimizing the uncertainty from the trend function of the model.

Kumar and Remadevi (2006) compared the OK method against a deterministic interpolation and inverse distance weighting (IDW) methods for the application of groundwater level interpolation in the northwestern part of Rajasthan, India. In their study, independent variograms were fitted for pre-monsoon and post-monsoon groundwater levels for six consecutive years and were interpolated by corresponding variogram model parameters. They found that the OK method outperforms the IDW method in the groundwater level interpolation process. A study by Ahmadi and Sedghamiz (2008) used the OK method to model the spatial distribution of groundwater levels in the southern parts of Iran. The study predicted groundwater level drops with reasonable accuracy at unvisited locations of the study area using cross-validation analysis. Furthermore, the study emphasized on identifying locations with severe water level drops on a spatial scale which is critical for water managers and land-use planners to optimally manage the groundwater resources and cropping systems, respectively. Chung and Rogers (2012) compared multiple

linear regression (MLR), OK and co-kriging (CK) methods for interpolating groundwater levels at St. Louis metro area, New Madrid seismic zone, St. Louis County. The MLR coefficients corresponding to minimum water table elevation and minimum water table depth were estimated against the dependent variable of water table elevation data. The MLR method closely mimics the groundwater table depths under the highly undulating topographic surface than CK, whereas OK is little influenced under the same conditions. Kumar and Ahmed (2003) applied the UK method with a linear drift to interpolate groundwater levels in Maheshwaram watershed located in Andhra Pradesh, India. Simple linear and quadratic trend equations with space coordinates were fitted to calculate the drift component of UK. The residual variance variogram can easily be modeled with less uncertainty when groundwater level data show a linear drift with space coordinates. In another similar study by Adhikary and Dash (2017), the pre-monsoon and post-monsoon groundwater table depths from 116 locations over the space were interpolated using IDW, OK, UK, and radial basis function models. They reported that the UK method outperformed over OK and IDW methods. Various authors analyzed the accuracy in spatial modeling of groundwater levels by comparing different kriging methods with model performance indices such as mean error (ME), mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) (Kumar 2007; Nikroo et al. 2010; Delbari et al. 2013; Adhikary and Dash 2017).

The RK models are useful in predicting groundwater levels when the trend function in RK captures most of the variance in the groundwater level datasets. More often the trend function is fitted between groundwater levels and the variable that is strongly associated with the groundwater levels. For example, the land surface elevations or DEM are often strongly correlated with the shallow unconfined aquifer groundwater levels. Therefore, the DEM is routinely used as a main auxiliary variable in RK methods for the prediction of groundwater levels in the spatial domain. The advantages of DEM-based trend surface modeling in groundwater level interpolation were demonstrated with a linear and two quadratic functions with and without considering DEM variables in the Fuyang River Basin, North China, by Zhu et al. (2013). Furthermore, the study justified that correlating DEM as an auxiliary variable with groundwater levels was more appropriate in predicting the groundwater levels as there was a strong dependency between groundwater levels at 23 observation wells and corresponding DEM values with a coefficient of determination (R^2) of 0.95. However, the study also reported that the discrepancy in considering only DEM values in one of the trend surface functions generated uneven distribution of residuals from the trend models. This fact emphasizes

that the DEM variable alone is not perfectly filtering the spatial trend from the groundwater level datasets.

A study by Desbarats et al. (2002) used collateral surface elevation information as external drifts in groundwater level predictions where the trend was modeled with a linear deterministic function given by TOPMODEL topographic index. The trend surface function with standalone DEM and groundwater levels was developed as another kriging-based external drift (KED) model. The benchmark OK model without the surface collateral information was developed to compare with the other two KED models. The cross-validation statistics for MSE show that those values were slightly more than the nugget variance of the experimental variograms which indicates that most of the error in the cross-validation statistics attributed to the noise in the data due to temporal fluctuations in the water levels. The study reported that the largest prediction error was observed in the deepest water table locations in the study area. It further emphasizes that the groundwater level predictions based on collateral surface elevation information alone may not be able to capture the maximum variability in the predicted groundwater levels especially when the groundwater level dataset shows a considerable variation in the water levels.

The spatial prediction performance of the groundwater levels by standard kriging-based methods including the RK method shows a relatively better result over physically based IDW methods where the distribution of observation points is relatively uniform and dense (Kumar 2007). However, the spatial prediction of the groundwater levels under sparsely distributed datasets is also vital as the measurements in the remote locations are often not available in many parts of the world. A study by Varouchakis and Hristopulos (2013) attempted to improve the spatial prediction of the groundwater levels with sparsely distributed datasets in the island of Crete, Greece, using RK method in which they used auxiliary trend variables based on DS and physical-based Thiem's multiple-well equations in the trend functions to predict the groundwater levels. The study compared the prediction model results across the OK, UK, DS, DEM, and Thiem's equation-based RK models. The DS, DEM, and Thiem's equation-based RK models were relatively superior over OK and UK methods. In addition, the study reported that the estimated groundwater levels in the far-away location from the measurement points were predicted with relatively higher uncertainty although the RK model trend functions were reducing the residual variance to a significant extent. The estimated standard deviations in the predicted groundwater levels were varied up to 4.5 m above mean sea level (MSL). It indicates that the improvement in the groundwater level predictions is still possible with a novel trend function which captures relatively higher variability from sparse distribution of groundwater level observations.

In this background, the present study aims to introduce a mean groundwater level variable as an auxiliary variable in the trend function as a part of the RK model to improve the overall groundwater level predictions under sparsely distributed observations. The study proposes a new trend function with mean groundwater level generated using long term groundwater level datasets from the well locations used in the RK approach, to effectively capture the trend component in sparsely distributed well locations, thus minimizing the overall uncertainty in groundwater level predictions. The study hypothesizes that the mean value of the groundwater levels can be considered as a drift variable instead of a routinely used standalone DEM as a drift variable as the mean groundwater levels are highly correlated with groundwater levels at a specific time than DEM variable. Moreover, the mean groundwater levels-based interpolation methods can be applied to all ranges of groundwater hydraulic heads or depth to water table data (shallow, medium and deep water levels), unlike DEM-based interpolation methods which are most suitable for shallow groundwater level datasets (Desbarats et al. 2002). The present study uses the unique explanatory variable, mean groundwater level, as a part of a trend function in the RK model to predict groundwater levels. The specific objectives of the present study are: (1) to develop a new trend function with mean groundwater level as an auxiliary variable in RK method and (2) to assess the performance of the new trend function-based RK method with a deterministic IDW method and other kriging-based methods such as OK and RK methods with standalone DEM and DS as the auxiliary variables in the trend functions.

Methodology

Inverse distance weighting method

The basic concept of the IDW method is that the closest points are more relevant in predicting a better value than farther away points. The IDW method uses the weight function, which is inversely related to distances from the prediction point to the known surrounding points. Furthermore, the distances are raised to the power value of p. The value of p plays a major role in the final predicted values. As the value of p increases, the weights for the distant points decrease rapidly. The formulation of the IDW is as follows:

$$\hat{z}(x_0) = \frac{\sum_{x=1}^{N} \frac{z_x}{d_x^p}}{\sum_{x=1}^{N} \frac{1}{d_x^p}} \qquad (1)$$

where $\hat{z}(x_0)$ is the predicted value at an unknown location x_0 (L); z_x is the value of the sample from the prediction location x_0 (L); d_x is the distance from the prediction location to

the surrounding sample points locations (L); p is the power function; and N is the number of sample points.

Spatial modeling of groundwater levels by kriging techniques

Semivariogram analysis

A semivariogram can model the spatial structure of groundwater level. A semivariogram is a functional relationship between the sampled variance calculated from observed data pairs and distance separated by a lag distance h. The semivariance can be estimated as follows (Isaaks and Srivastava 1989):

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

where $\gamma(h_{lag})$ is estimated value of the semivariance at lag h_{lag} [L^2]; $N(h_{lag})$ is the number of experimental data pairs separated by a lag distance h_{lag}; $Z(x_i)$ and $Z(x_i + h)$ are the values of the variables at locations x_i and $x_i + h_{lag}$ respectively [L].

The experimental variogram data points can be fitted by many theoretical models such as spherical, exponential, and Gaussian models. However, the spherical model is the most efficient theoretical model for fitting experimental data for modeling groundwater level as compared to exponential and Gaussian models (Ma et al. 1999). Therefore, the spherical model was adopted in the present study to model the spatial structure of the groundwater levels. The spherical model formulation is given as follows (Clark 1979):

$$\gamma(h_{lag}) = \begin{cases} C_0 + C_1 \left[\frac{3}{2} \frac{h_{lag}}{a} - \frac{1}{2} \left(\frac{h_{lag}}{a} \right)^3 \right] & h_{lag} \leq a \\ C_0 + C_1 & h_{lag} > a \end{cases} \quad (3)$$

where C_0 is the semivariogram intercept in the y-axis (nugget effect) [L^2]; $C_0 + C_1$ is sill variance of the semivariogram [L^2]; and a is a range of the semivariogram [L].

The sill is the variance value at which semi-variance cloud points level off in the semivariogram plot. In other words, it is the maximum variance value to be modeled in the semivariogram analysis. The range is the lag distance up to which the autocorrelation process exists significantly and beyond which the autocorrelation decreases to zero. All the models, including IDW, kriging, and semivariogram models, were coded in MATLAB software (MATLAB 2014a version) in this study.

Ordinary kriging

Different kinds of kriging methods can be applied for predicting groundwater levels in the spatial domain depending on the nature of the groundwater level data and availability of additional auxiliary variable datasets such as DEM. However, the OK method uses only the variable of interest which, in this case, is groundwater level data. The OK model formulation is given below:

$$Z(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) \quad (4)$$

$$\begin{cases} \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_j) - \mu = \gamma(x_i, x_0) \\ \sum_{i=1}^{N} \lambda_i = 1 \end{cases} \quad (5)$$

where $Z(x_0)$ is unbiased estimator of the prediction variable at the unvisited location x_0 [L]; λ_i is the kriging weights between the unvisited location x_0 and the visited location x_i where $i = 1, 2, 3 \ldots N$ are the number of observation locations; μ is the Lagrange coefficient; $\gamma(x_i, x_j)$ is average semivariogram values between the location x_i and extremity location x_j [L^2]; and $\gamma(x_i, x_0)$ is average semivariogram value between the location x_i and the prediction location x_0 [L^2].

The important advantage of the kriging methods over deterministic methods is that it can estimate the prediction error variance along with the prediction results. The prediction error variance of the OK method in a matrix form can be written as follows:

$$\sigma^2_{OK}(x_0) = (C_0 + C_1) - c_0^T . C^{-1} . c_0 \quad (6)$$

$$c_0 = \{C(x_0, x_1), ..., C(x_0, x_N)\}^T \quad (7)$$

$$C = \begin{bmatrix} C(x_1, x_1) & ... & C(x_1, x_N) \\ . & . & . \\ C(x_1, x_N) & ... & C(x_N, x_N) \end{bmatrix} \quad (8)$$

where $\sigma^2_{OK}(x_0)$ is the OK variance of the estimation error at x_0; c_0 is covariance vector at the unvisited locations; and C is a covariance matrix.

Regression kriging

The OK method assumes that the groundwater level data is stationary, and the trend is a constant, which is also an unknown process over the spatial domain. However, on many occasions, groundwater level data is non-stationary, and it shows a robust spatial trend with other attributes such

as spatial coordinates and elevation values. This kind of deterministic spatial trend can be modeled separately along with the stochastic variation of the groundwater level data as a residual component using the RK approach. The general formulation of RK is given as follows (Hengl et al. 2003; Varouchakis and Hristopulos 2013):

$$\hat{Z}(x) = \hat{m}(x) + \hat{\varepsilon}(x) \tag{9}$$

where $\hat{Z}(x)$ is predicted water level at location x [L]; $\hat{m}(x)$ is fitted trend value at location x [L]; and $\hat{\varepsilon}(x)$ is trend surface residual at location x [L].

Trend function with distance to stream network data

The correlation among the groundwater level surface and distance from the stream network is often studied as the groundwater aquifer, and surface water systems interact with each other. Therefore, in this study, we developed a simple linear function between groundwater levels and distance to the stream network variable. As a first step in generating the stream network map from DEM, the flow accumulation raster map was generated from a sink-filled DEM data. The flow accumulation map was used to create the streams network map in Adyar River Basin. We used the SAGA GIS module as a part of the QGIS package (https://qgis.osgeo.org) to create the GIS layers such as flow accumulation raster maps and stream network vector layers. The stream network layer and the vectorized DEM data were used to create the closest distance from the stream network raster map using distance to the nearest hub module in QGIS. A simple linear function was fitted between the groundwater level data from the well locations and their corresponding distance to stream network values. The formulation of the trend function between the groundwater level and distance to stream network (RK-DS) is as follows:

$$m(x)_{DS} = k_{DS} + a_{DS}(\text{DS}) + \varepsilon \tag{10}$$

where $m(x)_{DS}$ is a deterministic trend value of groundwater hydraulic heads based on DS variable [L]; a_{DS}, k_{DS} are least-square fitting parameters; and ε is the trend residual assumed to be normally distributed [L] with mean 0 and variance σ^2.

Trend function with standalone DEM data

The shallow unconfined aquifer systems groundwater level surface often reflects the surface topography (Desbarats et al. 2002). As there is a strong correlation between groundwater levels and surface elevation values, the DEM values are accounted as an auxiliary variable in the trend model of the RK method. Sometimes, the trend function can also be drifted with DEM along with spatial coordinates (Kumar and Ahmed 2003; Gundogdu and Guney 2007; Zhu et al. 2013). In the

present study, a simple linear function between groundwater level data (dependent variable) and DEM values (independent variables) (RK-DEM) can be fitted as follows:

$$m(x)_{DEM} = k_{DEM} + a_{DEM}(\text{DEM}) + \varepsilon \tag{11}$$

where $m(x)_{DEM}$ is a deterministic trend value based on DEM [L]; a_{DEM}, k_{DEM} are least-square fitting parameters; and ε is the trend residual assumed to be normally distributed [L] with mean 0 and variance σ^2.

The parameters of Eq. (11) were obtained by the least square method of fitting a function.

Proposed trend function with mean groundwater level as a function of DEM

In the present study, a new trend function was proposed with a long-term mean groundwater level (MGWL) variable as a function of the predicted groundwater levels. The schematic diagram of the proposed trend function is shown in Fig. 1.

The variable, MGWL, was calculated at the observation well locations using a simple arithmetic averaging method from the existing long-term monthly groundwater level datasets. As the groundwater level variation is influenced by the season, MGWL values were calculated based on long-term monthly average groundwater levels (for January through December) at the observation well locations (Fig. 1). This process estimates only point-based MGWL values at the well locations for all the months (i.e., January through December). However, the spatially continuous grid-based MGWL raster maps (like DEM) for the corresponding months are needed to model the spatial trend from the corresponding month's groundwater level data over the entire spatial domain. This demands the development of a spatially continuous or gridded MGWL maps for all the months in the study area. The development of the gridded MGWL maps from DEM grid values is explained in two steps as follows:

> Step 1: Fitting a linear function for MGWL and DEM at well sites.

First, a simple linear function was established between MGWL and DEM (MGWL-DEM) values based on well sites data for the corresponding months as follows:

$$\sum_{mi=1}^{12} \sum_{i=1}^{N} \text{MGWL}_{i,mi} = \sum_{mi=1}^{12} ks_{mi} + \sum_{mi=1}^{12} cs_{mi} \left[\sum_{i=1}^{N} \text{DEM}_i \right] + \varepsilon \tag{12}$$

where $\text{MGWL}_{i,mi}$ is the mean groundwater levels at the well locations $i = 1...N$ and for the months $mi =$ January through December [L]; DEM_i is DEM values at the corresponding well locations $i = 1...N$; ks_{mi} and cs_{mi} are generalized least-square fitting parameters of linear MGWL-DEM model for
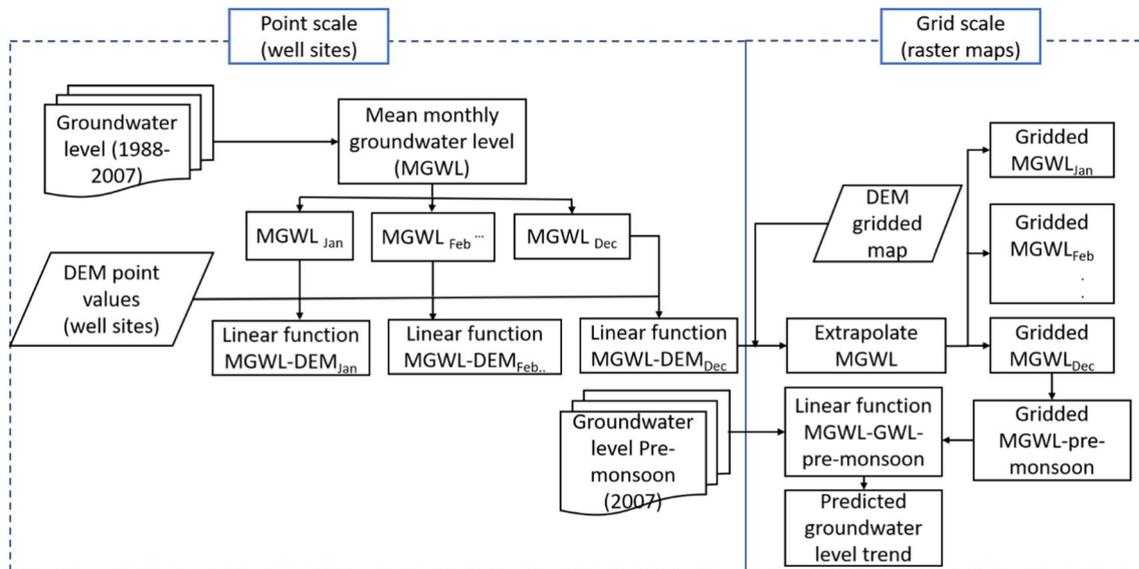
Fig. 1 The proposed trend model based on mean groundwater level as an auxiliary variable

the months $mi =$ January through December [L]; and N is number of observation well sites.

The coefficients of Eq. (12) were estimated using a generalized least square method of minimizing the errors. As a functional relationship between MGWL and DEM is established through Eq. (12), it can be used to estimate the gridded MGWL maps for the months January through December by extrapolation of gridded DEM values.

Step 2: Developing a new trend function with MGWL as a predictor variable in RK trend model.

A new trend equation based on spatially derived MGWL variable (Eq. 12), RK-MGWL, is expressed as follows:

$$m(x)_{MGWL} = k_{mi} + b_{mi}(MGWL_{mi}) + \varepsilon \tag{13}$$

where $m(x)_{MGWL}$ is a deterministic trend value filtered from groundwater level based on MGWL values [L]; k_{mi}, b_{mi} are fitting parameters; and ε is the trend residual assumed to be normally distributed [L] with mean 0 and variance σ^2.

The spatially continuous gridded trend surface was generated by extrapolating the values at the grid points with the corresponding gridded MGWL values using Eq. (13). It is assumed that the calibrated parameters k_{mi}, b_{mi} based on well site data are the representative values for the whole study region of interest. Thus, they can be used for predicting the trend values at all unvisited grid locations in the study area. The same assumptions were applied for the other trend models such as RK-DS and RK-DEM (Eqs. 10 and 11) functions for generating the gridded trend surface maps.

The residual part of RK

The spatial trend in the groundwater level data can be filtered by the deterministic trend functions using Eqs. (10), (11) and, (13). The trend component of the RK model was calculated after estimating the parameters for Eqs. (10), (11) and, (13). Once the trend component from the groundwater level data was determined, the residual component was calculated as the difference between the observed groundwater hydraulic head values and the calculated linear trend model values (Eqs. 10, 11, and 13) at well sites. These estimated residuals were then modeled over the entire spatial domain by adopting the OK method (similar to Eqns. 4 and 5) as follows:

$$\varepsilon(x_0) = \sum_{i=1}^{N} \lambda_i \varepsilon(x_i) \tag{14}$$

$$\begin{cases} \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_j) - \mu = \gamma(x_i, x_0) \\ \sum_{i=1}^{N} \lambda_i = 1 \end{cases} \tag{15}$$

Prediction equations and error variance of RK

The estimated deterministic gridded trend component and the interpolated stochastic residual component were added together to get the total predicted gridded groundwater levels

according to Eq. (9). Therefore, the final prediction equations for RK-SD, RK-DEM, and RK-MGWL models can be written based on Eq. (9) as follows:

$$\hat{Z}_{RK-SD}(x_0) = \hat{k}_{SD} + \hat{a}_{SD}(\text{SD}) + \sum_{i=1}^{N} \lambda_i \hat{\varepsilon}(x_i) \tag{16}$$

$$\hat{Z}_{RK-DEM}(x_0) = \hat{k}_{DEM} + \hat{a}_{DEM}(\text{DEM}) + \sum_{i=1}^{N} \lambda_i \hat{\varepsilon}(x_i) \tag{17}$$

$$\hat{Z}_{RK-MGWL}(x_0) = \hat{k}_{mi} + \hat{b}_{mi}(\text{MGWL}_{mi}) + \sum_{i=1}^{N} \lambda_i \hat{\varepsilon}(x_i) \tag{18}$$

In general, the prediction error variance of RK model can be written as follows (Hengl et al. 2003; Varouchakis and Hristopulos 2013):

$$\sigma_{RK}^2(x_0) = \sigma_{RK}^2\{\hat{m}(x_0)\} + \sigma_{RK}^2\{\hat{\varepsilon}(x_0)\} \tag{19}$$

For a clear understanding of the terminologies, furthermore, the kriging variance of RK model in a matrix form can be calculated as follows (Hengl et al. 2003; Varouchakis and Hristopulos 2013):

$$\sigma_{RK}^2(x_0) = \{q_0^T \cdot (q^T \cdot C^{-1} \cdot q)^{-1} \cdot q_0\} + \{(C_0 + C_1) - c_0^T \cdot C^{-1} \cdot c_0\} \tag{20}$$

where q is the vector of the number of independent variables used at well sites (visited) locations; and q_0 is the vector of the number of independent variables used at unknown (unvisited) locations x_0.

Model evaluation methods and performance indices

Cross-validation is a powerful technique to evaluate the model predictions over a given spatial domain (Cooper and Istok 1988; Isaaks and Srivastava 1989). Leave one out validation (LOOV) is the widely used method in the cross-validation technique to effectively assess the model performance based on the observations. The LOOV method was carried out by first leaving an observation intentionally at a place and modeling with the remaining observation points. This process was repeated for all observations sequentially one after another. In each iteration, the predicted values at the locations where the observations were not accounted were compared with the corresponding observed values from the same locations. The model performance indices such as ME, MAE, MSE, and RMSE are formulated as follows (Delbari et al. 2013):

$$ME = \frac{1}{N} \sum_{i=1}^{N} Z(\hat{x}_i) - Z(x_i) \tag{21}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |Z(\hat{x}_i) - Z(x_i)| \tag{22}$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left[Z(\hat{x}_i) - Z(x_i) \right]^2 \tag{23}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[(Z(\hat{x}_i) - Z(x_i)) \right]^2} \tag{24}$$

$$R^2 = 1 - \left(\frac{\sum_{t=1}^{N} (Z(x_i) - Z(\hat{x}_i))}{\sum_{t=1}^{N} (Z(x_i) - Z(\overline{x}_i))} \right)^2 \tag{25}$$

where $Z(\hat{x}_i)$ is estimated value of the variable [L]; $Z(x_i)$ is observed value of the variable [L]; $Z(\overline{x}_i)$ is mean value of the observations; and N is total number of observations.

Case study: Geography of Adyar basin

For testing the effectiveness of the developed IDW and kriging methods, a sub-catchment of the Chennai Basin, Adyar basin, was chosen in the current study. The Adyar basin is geographically located in the North East coastal part of Tamil Nadu (Fig. 1). The Adyar River is one of the major rivers in the Chennai Basin which drains rainwater to the Bay of Bengal Sea during monsoon season. Semiarid and humid climatic zones characterize the prevailing climate in the basin. The long-term annual average rainfall in the basin is about 1315 mm (WRO 2007). The elevation ranges from 183 m amsl in the western part of the basin to sea level in the eastern part of the basin.

The soils in the basin have been classified into clayey, black, red sandy, and alluvial soils. The black soils occur in the depressions adjacent to hilly areas in the western part. The alluvial soils occur along with the river courses and eastern part of the coastal areas. The major hydrogeology in the basin has been classified as unconsolidated, semi-consolidated, and weathered fractured rock formation. The groundwater occurs under phreatic and semi-confined conditions in inter-granular pore spaces in sands, sandstones, bedding planes, and thin fractures in shales. The depth to groundwater table fluctuation in the observation wells varies minimum from near-surface to the maximum of 12 m below ground surface. The distribution of observation wells is more concentrated in the eastern part and sparse in the western parts of the basin. Also, as the surface elevations vary from the eastern to the western part with irregular

topographical conditions, the occurrence and distribution of groundwater vary widely from eastern to the western parts in the basin. As a result, there are considerable variations in the observed groundwater levels across the basin.

Data used in the analysis

Monthly groundwater level data of 29 observation wells for the period 1988–2007 (20 years) were acquired from the Institute of Water Studies (IWS), Taramani, Tamil Nadu, India. The Shuttle Radar Topography Mission (SRTM)-based DEM gridded data (90 m spatial resolution) for the study area was downloaded from the SRTM-Earth Models website (https://srtm.csi.cgiar.org/SELECTION/inputCoord.asp). The depth to groundwater table data was converted to groundwater level elevation or groundwater hydraulic heads with respect to MSL by subtracting depth to water table values from SRTM-DEM values at the observation well sites.

Pre-monsoon and post-monsoon season's groundwater level data

For brevity, the pre-monsoon (June) and post-monsoon (December) seasons groundwater level datasets of the year 2007 were chosen for predicting a spatially continuous gridded groundwater level using the developed IDW and kriging methods. Typically, the groundwater level during pre-monsoon season is relatively deeper from the surface while the groundwater level rises close to the surface during post-monsoon seasons. As the variation in the water levels during these months are higher (June and December), these two months datasets were considered for demonstrating the performance of various developed models in this study.

Results and discussion

Groundwater hydraulic head data analysis

Most of the observation wells in the study area were concentrated in the eastern part of the basin where the groundwater levels are located at shallow depth and mostly consists of alluvial formation in the coastal region of the basin. Therefore, the actual groundwater level values with respect to MSL were also relatively lower. The groundwater hydraulic head values vary from near-surface to few meters below the surface in some of the observation wells, which are located closer to the coastal boundary of the basin (Fig. 3). On the contrary, the groundwater levels were typically varied up to 10 m in some of the observation wells located in the central part of the basin where hard rocks characterize the topography. For example, the well nos. 14 and 20 (Fig. 2) typically lie in the central part of the basin that belongs to hard rock

hilly regions showing a relatively higher variation in the groundwater levels (Fig. 3).

Furthermore, three representative wells (well nos. 8, 20, and 9) from eastern, central, and western parts of the study area and corresponding temporal variations in the groundwater levels are shown in Fig. 4. As the elevation decreases from the western to the eastern part in the study area, a clear distinction in the hydraulic head values was noticed in the representative wells (Fig. 4). For example, the hydraulic head values of the western part of the wells were ranging from 25–30 m amsl due to relatively higher topography of the region. On the other hand, the hydraulic head values of the eastern part show the variation between 5–10 m due to relatively flatter topography. A relatively higher variation in the groundwater hydraulic heads (12–22 m) was noticed for the central part of the well where the geological settings have been characterized as hard rock formations with rugged hilly terrain (Fig. 4). No significant trend was observed in any of the observed groundwater level data, including the three plotted groundwater level datasets shown in Fig. 4. Therefore, the proposed trend function, which is based on the MGWL function, is valid as the calculated long-term mean groundwater level is derived from these stationary groundwater level datasets.

Correlation of trend variables with groundwater hydraulic heads

The generated distance to stream map from SRTM DEM is shown in Fig. 5a. The distances from measurement sites to the nearest streams were ranging from 0 to 6250 m. The extreme distances from streams, more than 5000 m, were mostly located in the western and southern borders of the basin due to sparsely distributed measurement sites. Most of the observation wells, around 85 percent, fall within a stream network distance of 1250 m. However, few wells were located relatively away from the stream network (Fig. 5a). The calculated distance to stream values was extracted for all 29 observation wells. The correlation between the extracted distance to stream values and the groundwater hydraulic heads during pre-monsoon and post-monsoon seasons in 2007 is shown in Fig. 5b, c. It was found that the correlation was relatively weak between groundwater hydraulic heads and distance to stream values during both pre-monsoon and post-monsoon periods (Fig. 5b, c). This could be because the streams in the basin are not perennial streams; thus, the groundwater contribution to the streams is very limited during non-monsoon seasons as groundwater tables are at relatively higher depths. Moreover, due to heavy groundwater pumping to meet the water demands, the population and industrial activities in the region are causing a decrease in groundwater level. These factors, collectively, cause the surface streamflow systems disconnected from the groundwater

Fig. 2 Location map of the Adyar basin with observation well numbers
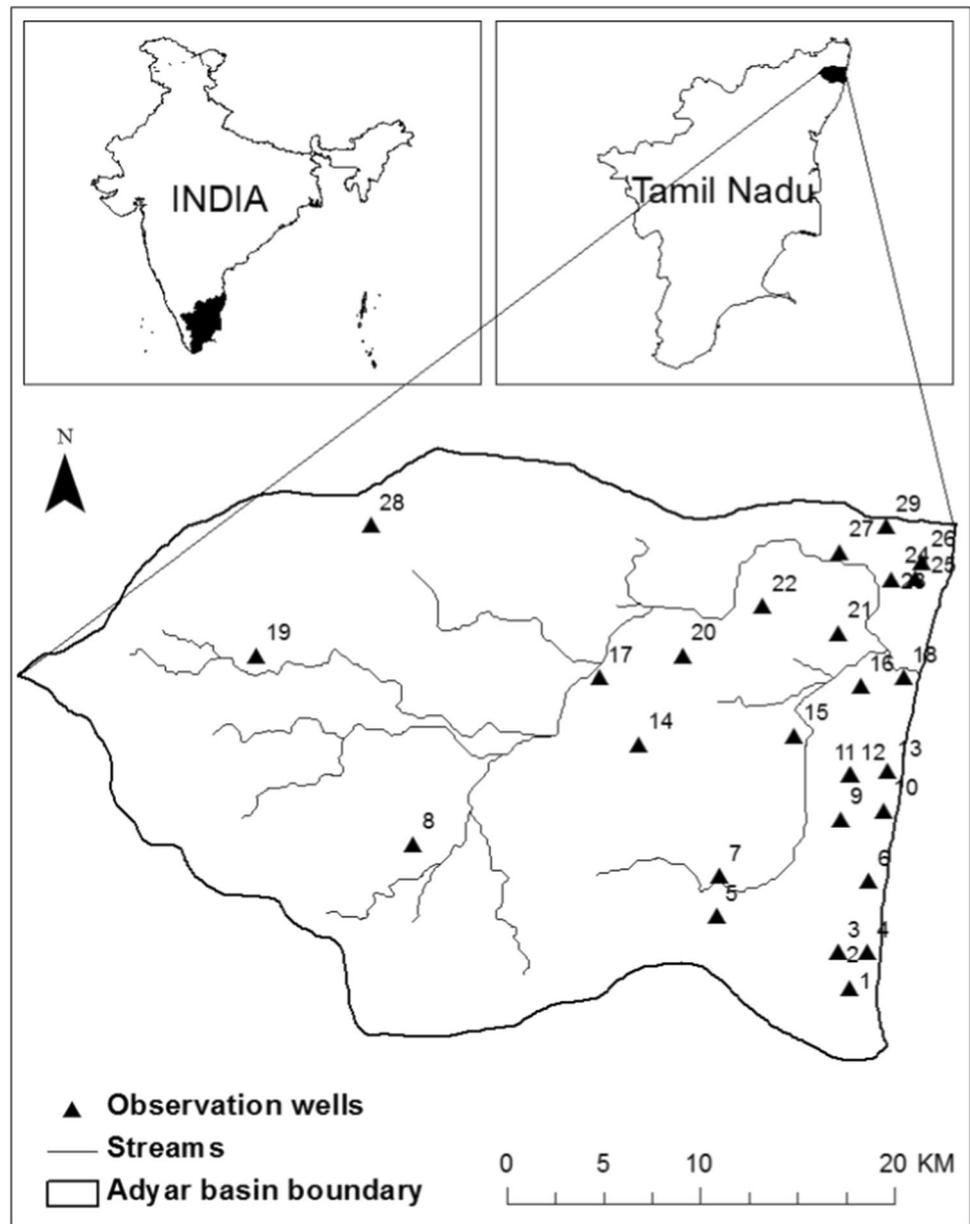
systems. Therefore, the groundwater hydraulic heads and distance to streams values were not highly correlated.

The scatter plots with linear relationship models of observed groundwater hydraulic heads against DEM and MGWL values at the well locations during pre-monsoon and post-monsoon seasons are shown in Fig. 6. Both DEM and MGWL values have a strong correlation with groundwater hydraulic heads during both the seasons. The DEM variable with groundwater hydraulic heads clearly shows a strong correlation as the topography of the land surface closely resembling with the shallow unconfined aquifer hydraulic head values in the Adyar basin (Fig. 6a, b). However, the MGWL variable with groundwater hydraulic heads shows a relatively higher correlation over DEM

variables with the correlation coefficient values of 0.99 and 0.99 during pre-monsoon and post-monsoon seasons respectively (Fig. 6c, d). This is because the MGWL values were calculated based on the long-term monthly average of groundwater levels and the correlation of any specific month groundwater levels with the monthly average values would naturally have a higher correlation. Therefore, we tested this hypothesis in this study that the MGWL variable will have a higher influence in separating the spatial trend from groundwater level datasets than any other variables, thereby the prediction accuracy would be improved in spatial modeling of groundwater hydraulic heads.

Fig. 3 Variation of groundwater hydraulic heads at the well sites in the study area during 1988–2007
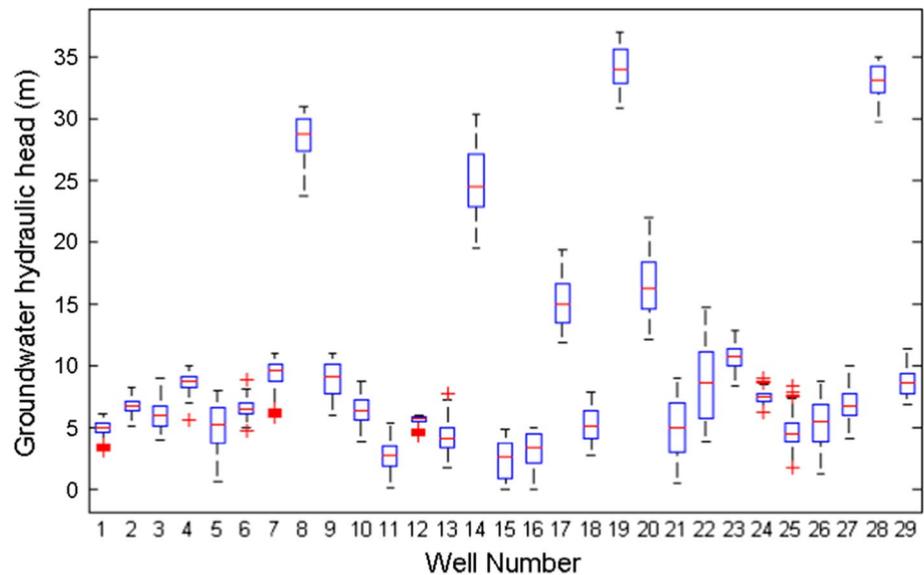
Estimation of trend model parameters at well sites

The parameters of the RK trend models (RK-DS, RK-DEM, and RK-MGWL) were estimated by the generalized least square (GLS) method with 29 observations wells datasets ($n = 29$). The t-statistics of the estimated model parameters are given in Table 1. It was observed that at least one of the two parameters of the linear models (slope and intercept) was statistically significant at 0.01 significance level (99% confidence limit) or 0.05 significance level (95% confidence limit) for all RK methods (Table 1). Although the distance to stream values with groundwater hydraulic heads was fairly correlated, the fitted linear regression model slope coefficient (k_{DS}) value was significantly different from zero at 0.01 significance level for both pre-monsoon and post-monsoon seasons. The slope parameters for both RK-DEM and RK-MGWL were significant at 0.01 significance level as DEM

and MGWL values were strongly correlated with groundwater hydraulic heads.

Unlike DEM, the MGWL is not a constant variable in temporal scale. As MGWL is a function of various hydrogeologic factors, it varies over both spatial and temporal dimensions. Therefore, it is also possible to establish unique functions for unique seasons or months if the observations are available for long time periods at the observation well sites. In present study, as the groundwater hydraulic head observations are available at a monthly interval for 20 years, twelve different DEM-MGWL functions with unique parameters for January through December were estimated by the GLS approach. The estimated parameters for Eq. (12) are given in Table 2. The estimated coefficients for DEM-MGWL models were statistically significant at 0.01 significance level for all the months as DEM and MGWL values were strongly correlated in the study area (Table 2). The spatially continuous MGWL maps over the study area were derived using the gridded DEM map as an independent variable and the optimized DEM-MGWL parameters for January through December. These 12 gridded MGWL maps were used for estimating trend components for corresponding months over the entire study domain in RK-MGWL trend models.

Pre-monsoon groundwater level predictions by IDW and kriging methods

The fitted semivariogram models of the developed kriging models using pre-monsoon groundwater level data are shown in Fig. 7. The parameter value for the number of bins or lags for calculating the experimental semivariance values was kept equal to 20. The spherical model coefficients values were initially assigned to arbitrary values, and later,

Fig. 4 Temporal variation of groundwater hydraulic head values at three unique locations in the basin

(a)

(b)

(c)
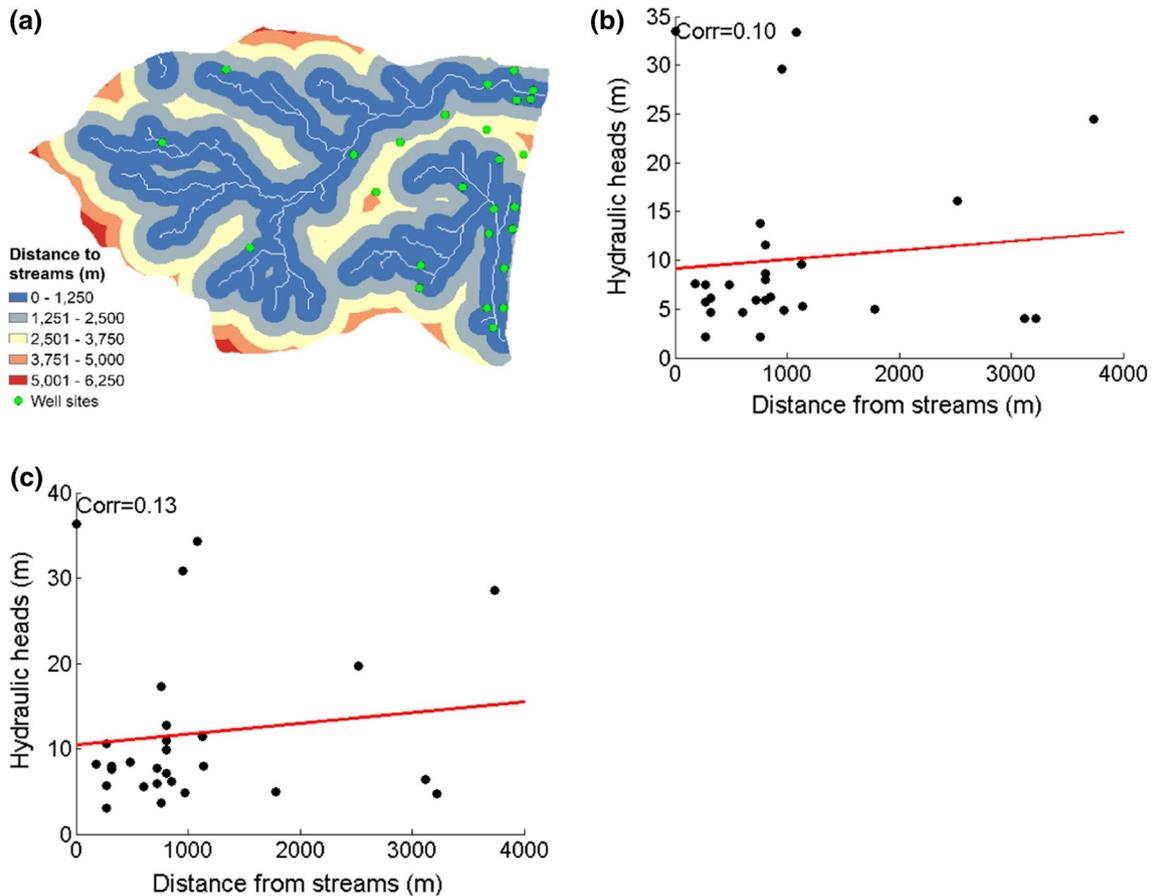
Fig. 5 **a** Distance to streams map and correlation of groundwater hydraulic heads and distance to streams data at the well sites during **b** pre-monsoon, and **c** post-monsoon season of the year 2007

it was optimized using 'fminsearch' optimization algorithm in the MATLAB software. The fitted semivariogram parameters, range and sill, for different kriging methods were different as the data points used to fit the semivariogram model were completely different depending upon the type of kriging model adopted in the present study (Fig. 7). The OK model used the actual groundwater hydraulic head data for constructing the semivariogram, whereas RK-DS, RK-DEM, and RK-MGWL methods used the residual data after removing the trend component from the actual groundwater hydraulic head signals to construct semivariograms. In all the cases, either the raw groundwater hydraulic heads or the trend removed residuals were checked for the normality conditions. The trend removed residuals data were normally distributed while raw actual groundwater hydraulic heads were approximately following the normal distribution. Therefore, data transformation to satisfy the normality conditions was not performed. The nugget effect in the semivariograms was observed in terms of a small semivariance value at a relatively a closer lag distance because two observations sampled at a closer distance not necessarily always have the

same value. Sometimes nugget effect could occur due to the measurement error in the observations.

The estimated range parameters for OK, RK-DS, and RK-DEM methods (17.69 km, 18.11 km, and 19.33 km) were relatively higher than the RK-MGWL method (6.82 km). This indicates that the spatial autocorrelation process quickly declines for the proposed RK-MGWL method than OK, RK-DS, and RK-DEM methods (Fig. 7). Similarly, the estimated sill values were 39.93 m^2 and 36.27 m^2 for OK and RK-DS models which were relatively higher than 5.14 m^2 and 1.35 m^2 for RK-DEM and RK-MGWL methods (Fig. 7). This sill variance for the OK method was relatively higher because of the higher variance in the observed groundwater hydraulic head data as the observation wells were sparsely distributed across the basin ranging from low altitude to high altitude regions in the basin (Fig. 7a). In general, the sill variance of the RK based models was relatively lesser as they used the residual data pairs for constructing the semivariograms. As the trend component in the RK model filters a major part of the variability in the groundwater level dataset, the remaining variability in residuals was relatively
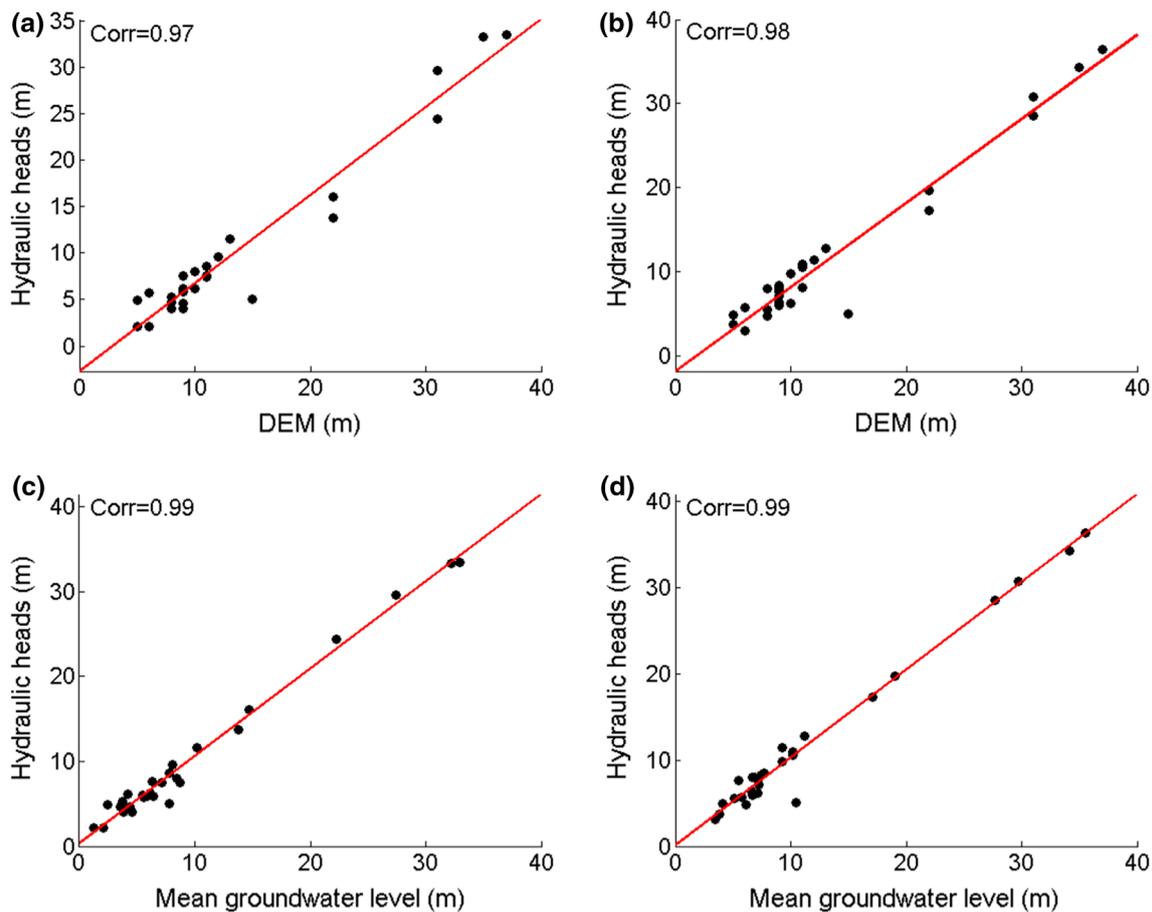
Fig. 6 Correlations between groundwater hydraulic heads and DEM at well sites during **a** pre-monsoon and **b** post-monsoon; and correlations between groundwater hydraulic heads and MGWL at well sites during **c** pre-monsoon and **d** post-monsoon seasons of the year 2007

Model	Pre-monsoon				Post-monsoon			
	Coefficient	Value	t value	p value	Coefficient	Value	t value	p value
RK-DS Equation (10)	k_{DS}	9.16	3.54	0.0015[**]	k_{DS}	10.41	3.85	6.9E−4[**]
	a_{DS}	9.2E−4	0.51	0.62	a_{DS}	0.0013	0.66	0.51
RK-DEM Equation (11)	k_{DEM}	−2.71	−3.69	0.001[**]	k_{DEM}	−1.80	−2.52	0.018[*]
	a_{DEM}	0.95	21.1	7.3E−18[**]	a_{DEM}	0.99	22.8	9.5E−19[**]
RK-MGWL Equation (13)	k_{mi}	0.3	0.98	0.33	k_{mi}	0.1	0.24	0.80
	b_{mi}	1.03	42.3	3.1E−26[**]	b_{mi}	1.01	35.6	3.0E−24[**]

Table 1 Estimated parameters of RK trend models for pre-monsoon and post-monsoon groundwater level datasets of the year 2007

[*] Significance at alpha level 0.05

[**] Significance at alpha level 0.01

smaller. This is the reason that RK-DEM and RK-MGWL methods show a relatively lower sill variance than the OK method (Fig. 7c, d). However, the RK-DS method shows a sill variance almost equal to the OK method, unlike other RK methods (36.27 m²) (Fig. 7b). This could be because of the poor performance of the trend function which captured only a small variability in the groundwater level data leaving major variability in the residual portion of the RK model

which was later modeled separately by the OK method. As the proposed trend surface function in RK-MGWL (Eq. 13) captures the maximum variability in the trend component of the RK model, the remaining residual variance values were minimal in magnitude. This was the reason for a significant reduction in the sill variance (1.35 m²) for the RK-MGWL method (Fig. 7d) as compared to other RK methods.

Table 2 Estimated parameters of DEM-MGWL relationship for the months January through December

Months indices (mi)	Coefficients	DEM-MGWL function (Eq. 12) Value		Months indices (mi)	Coefficients	DEM-MGWL function (Eq. 12) value	
		Value	p value			Value	p value
Jan	ks_{jan}	− 2.23	1.8E − 5[*]	Jul	ks_{jul}	− 2.71	4.4E − 5[*]
	cs_{jan}	0.97	1.9E − 24[*]		cs_{jul}	0.94	4.5E − 21[*]
Feb	ks_{feb}	− 1.92	3.7E − 5[*]	Aug	ks_{aug}	− 2.79	4.8E − 5[*]
	cs_{feb}	0.98	1.0E − 25[*]		cs_{aug}	0.92	1.5E − 20[*]
Mar	ks_{mar}	− 1.90	4.0E − 5[*]	Sep	ks_{sep}	− 2.86	4.7E − 5[*]
	cs_{mar}	0.99	7.9E − 26[*]		cs_{sep}	0.92	3.6E − 20[*]
Apr	ks_{apl}	− 2.17	1.6E − 5[*]	Oct	ks_{oct}	− 3.02	2.3E − 5[*]
	cs_{apl}	0.98	5.9E − 25[*]		cs_{oct}	0.92	3.2E − 20[*]
May	ks_{may}	− 2.45	1.5E − 5[*]	Nov	ks_{nov}	− 2.99	2.2E − 5[*]
	cs_{may}	0.96	1.8E − 23[*]		cs_{nov}	0.93	2.0E − 20[*]
Jun	ks_{jun}	− 2.62	3.0E − 5[*]	Dec	ks_{dec}	− 2.57	2.8E − 5[*]
	cs_{jun}	0.95	5.8E − 22[*]		cs_{dec}	0.94	3.7E − 22[*]

[*]Significant at alpha level 0.01
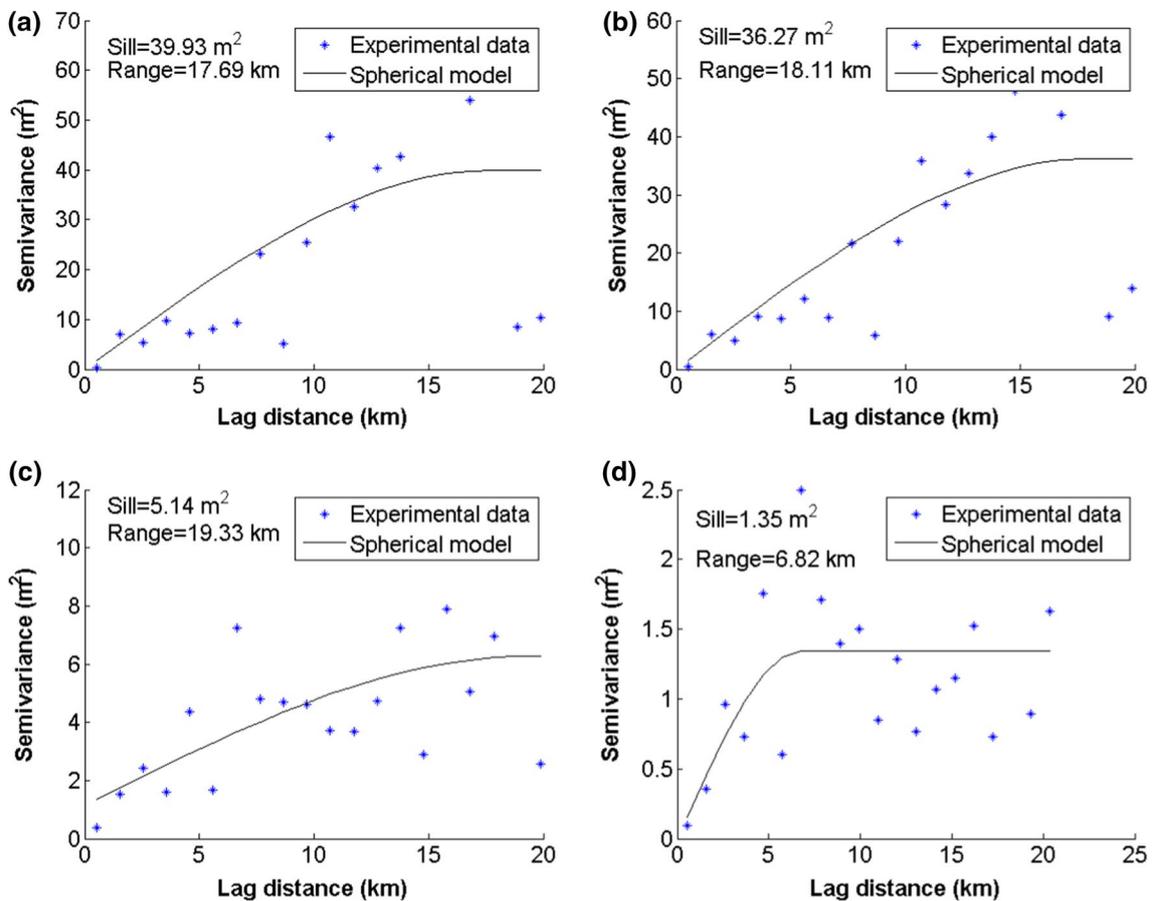
Fig. 7 Fitted spherical semivariogram models for **a** OK, **b** RK-DS, **c** RK-DEM, and **d** RK-MGWL using pre-monsoon season hydraulic head data for the year 2007

The total groundwater hydraulic head predictions by the developed kriging and IDW models are shown in Fig. 8. The intrinsic deterministic method of IDW interpolates groundwater hydraulic heads only based on the distance criteria. As there is no spatial covariance modeling accounted in the IDW method, it fails to produce a real variation in the groundwater levels under the sparse distribution of the measurement sites in the study area (Fig. 8a). As the OK

method (Eq. 4) could not capture the accurate spatial trend component in the high-altitude regions, it severely under predicts the groundwater levels in the western parts of the basin (Fig. 8b). For example, groundwater heads predicted by the OK method in the western parts were ranging from 30 to 35 m which were far away from the realistic groundwater hydraulic heads variation as the surface elevation in these regions varies about 50 to 70 m amsl (Fig. 8b). The distance
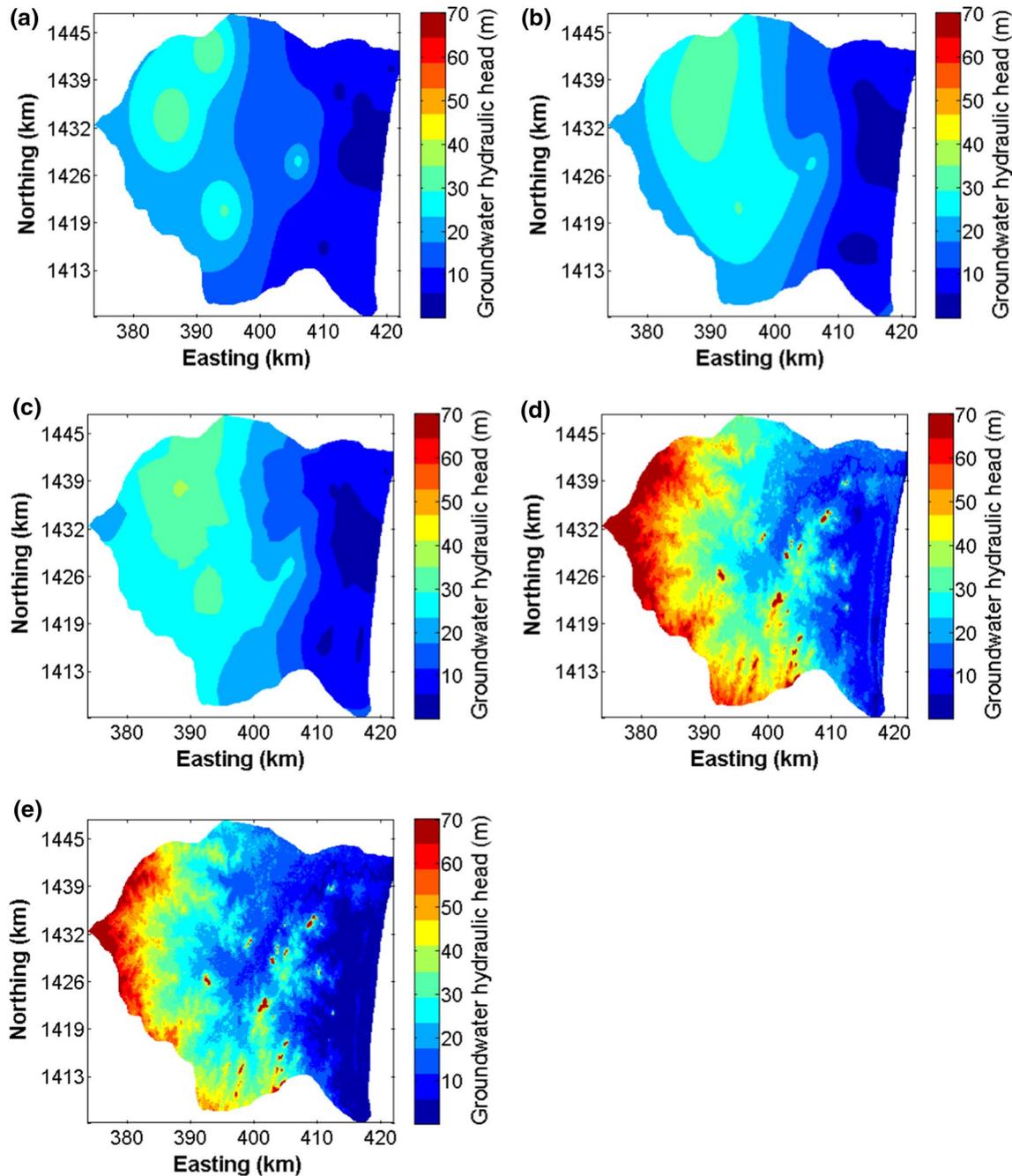
Fig. 8 Predicted groundwater hydraulic heads by **a** IDW, **b** OK, **c** RK-DS, **d** RK-DEM, and **e** RK-MGWL using pre-monsoon hydraulic head data

to stream parameter-based RK model produced a groundwater level surface map with the maximum hydraulic head value of 40–45 m in the western parts of the basin which are still underestimated values in that part of the basin (Fig. 8c). It was observed that the IDW, OK, and RK-DS methods estimate a relatively lower hydraulic head value even in the highly elevated regions of the basin. This is clear from Fig. 8a–c that the interpolated hydraulic head values near to the boundary were relatively lesser than in the central part of the basin. This leads to the conclusion that the groundwater flow moves from the central part to the surrounding boundary regions of the basin, which is not realistic based on the topography of the basin.

Interestingly, both RK-DEM and RK-MGWL methods (Eqs. (17) and (18)) which compute the trend component of the groundwater level data with respect to DEM and MGWL variables, respectively, predict a reasonably realistic variation of groundwater hydraulic heads across the basin (Fig. 8d, e). Although the groundwater hydraulic head prediction by both RK-DEM and RK-MGWL methods shows a similar pattern, there is a distinct variation among these two models' predictions (Fig. 8d, e). The groundwater hydraulic head predictions by both RK models (RK-DEM and

RK-MGWL) were ranging from few meters in the eastern parts to 70 m in the western parts of the basin which closely aligns with the existing topographic elevations of the basin. Unlike IDW, OK, and RK-DS methods, RK-DEM and RK-MGWL methods predict a realistic variation of groundwater hydraulic heads in the central hard rock hilly regions of the basin as well (Fig. 7d, e).

The total prediction error variance by OK (Eq. 6), RK-DS (Eq. 20), RK-DEM (Eq. 20), and RK-MGWL (Eq. 20) methods is shown in Fig. 9. The color code scheme of the total prediction variance maps is given with a range of 0 to 6 m^2 for RK-DEM and RK-MGWL methods and 0 to 30 m^2 for OK and RK-DS methods (Fig. 9). We did not provide the same color code scale for all the methods because the magnitude of the prediction error variance for OK and RK-DS methods was significantly higher than RK-DEM and RK-MGWL methods. Thus, it could potentially suppress the higher values variation in the OK and RK-DS-based total error variance maps. The prediction error variance was significantly higher (more than 50 m^2) for the OK method as the trend component was not modeled properly in the OK method (Fig. 9a). The same is the case for the RK-DS method (Fig. 9b), but the total prediction error variance was
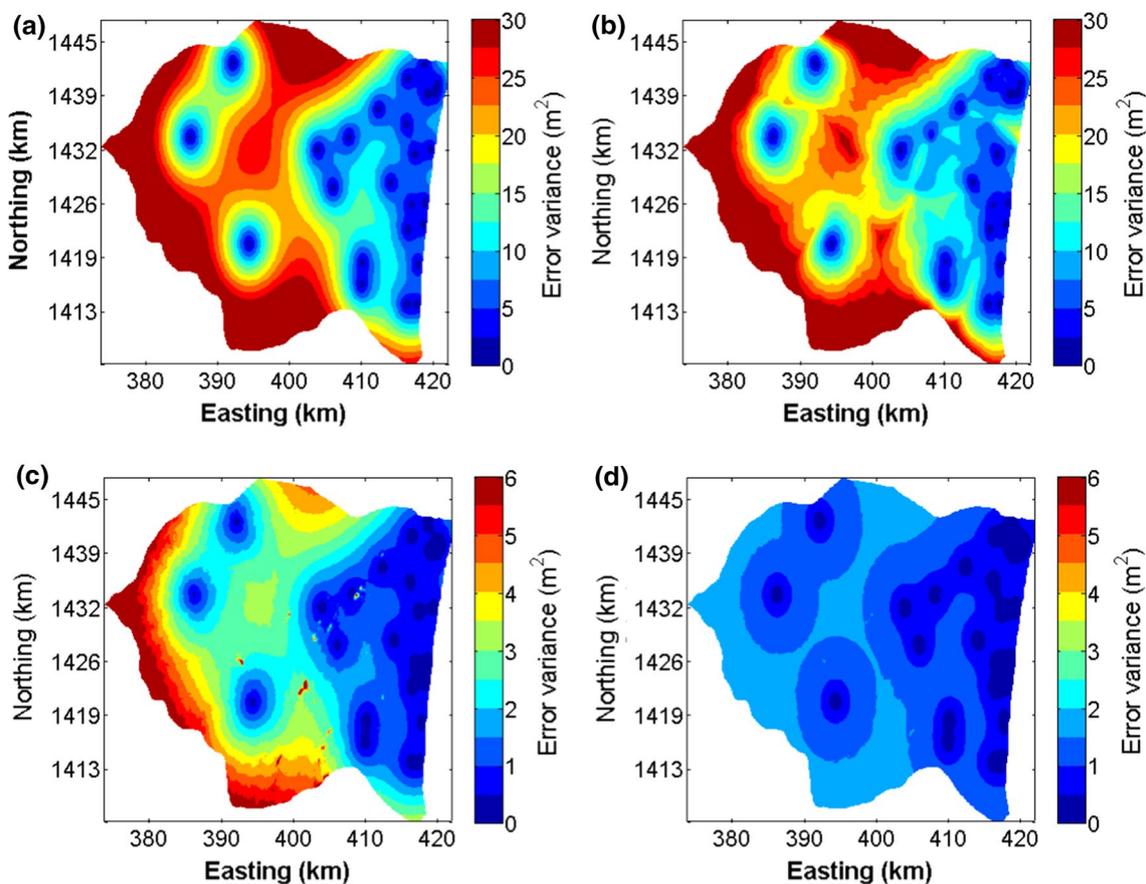
Fig. 9 Total error variance created by **a** OK, **b** RK-DS, **c** RK-DEM, and **d** RK-MGWL models using pre-monsoon hydraulic head data

significantly reduced for the RK-DEM method which was ranging from 0 to 6 m^2 (Fig. 9c). In general, it was observed that the total prediction error was comparatively higher up to 6 m^2 in the western parts of the basin due to the sparse distribution of observation wells in these regions. On the other hand, due to the dense availability of the observation wells in the central and eastern parts of the basin, the total prediction error was significantly reduced to less than 1 m^2 in these regions (Fig. 9c). Interestingly, RK-MGWL method was outperforming over OK, RK-DS, and RK-DEM models as the maximum prediction error itself was observed less than 1 m^2 over the entire basin (Fig. 9d). The prediction error variance was substantially reduced for this proposed method because the maximum variance in the groundwater hydraulic head data was effectively captured by the RK-MGWL trend function (Eq. 13).

Post-monsoon groundwater level predictions by kriging methods

The fitted semivariogram spherical models for the post-monsoon season groundwater level analysis were similar to

pre-monsoon season groundwater level analysis (Fig. 10). Therefore, as expected, the sill variance was higher for the OK method (50.41 m^2) followed by RK-DS (44.11 m^2), RK-DEM (5.21 m^2), and RK-MGWL (2.26 m^2) methods (Fig. 10).

Similarly, the groundwater hydraulic head prediction surfaces obtained during the post-monsoon season were similar to pre-monsoon season groundwater hydraulic head predictions (Fig. 11). The IDW, OK, and RK-DS methods predicted a relatively higher (40 m to 45 m) groundwater hydraulic head values in the northwestern part of the basin. In contrast, the remaining parts of the basin were predicted with relatively lower hydraulic head values (0 m to 30 m) (Fig. 11a–c). The RK-DEM and RK-MGWL methods predicted the groundwater hydraulic head values in a realistic manner such that the groundwater flow moves from relatively higher elevated regions in the western part of the basin towards the lower elevation in the eastern coastal boundary of the basin (Fig. 11d, e). In general, it was observed that the post-monsoon season groundwater hydraulic head predictions were relatively higher as compared to pre-monsoon season because the
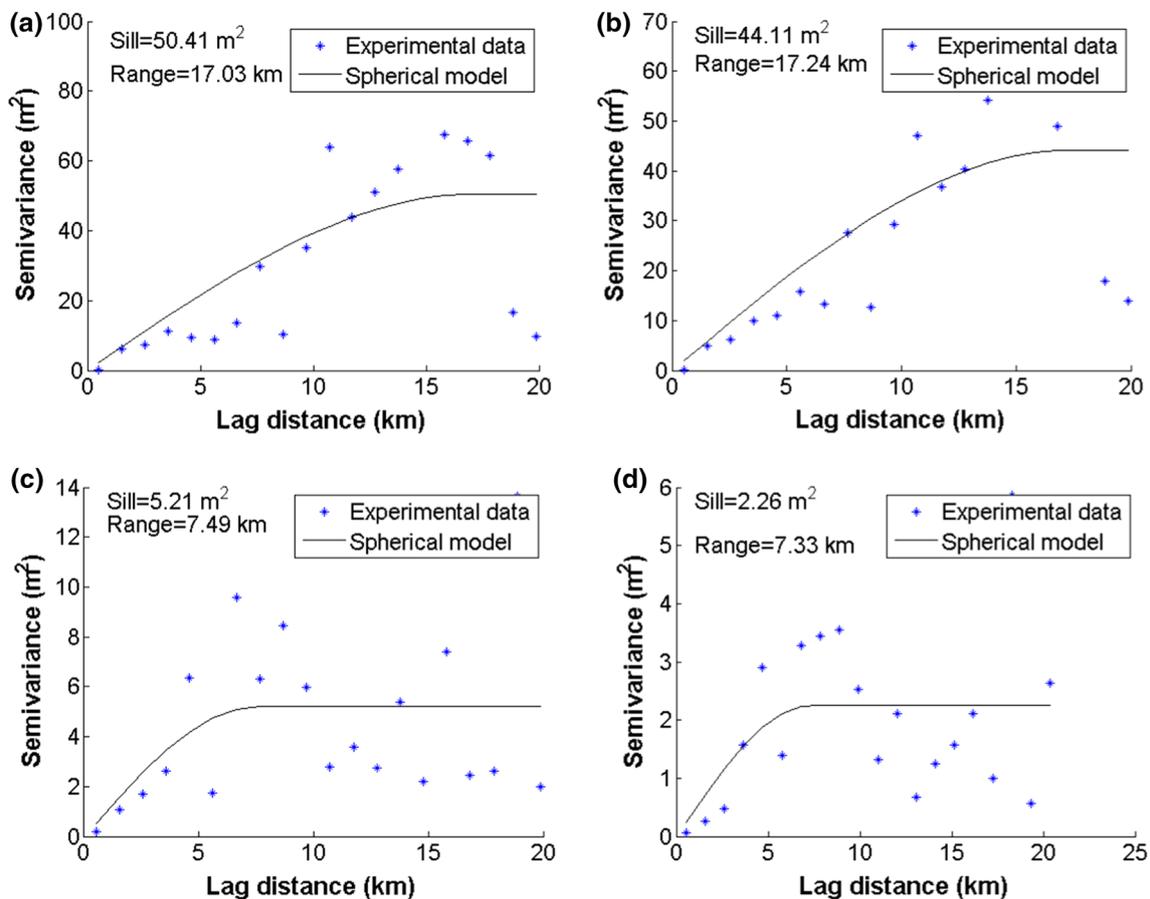
Fig. 10 Fitted semivariogram models for **a** OK, **b** RK-DS, **c** RK-DEM, and **d** RK-MGWL using post-monsoon season hydraulic head data
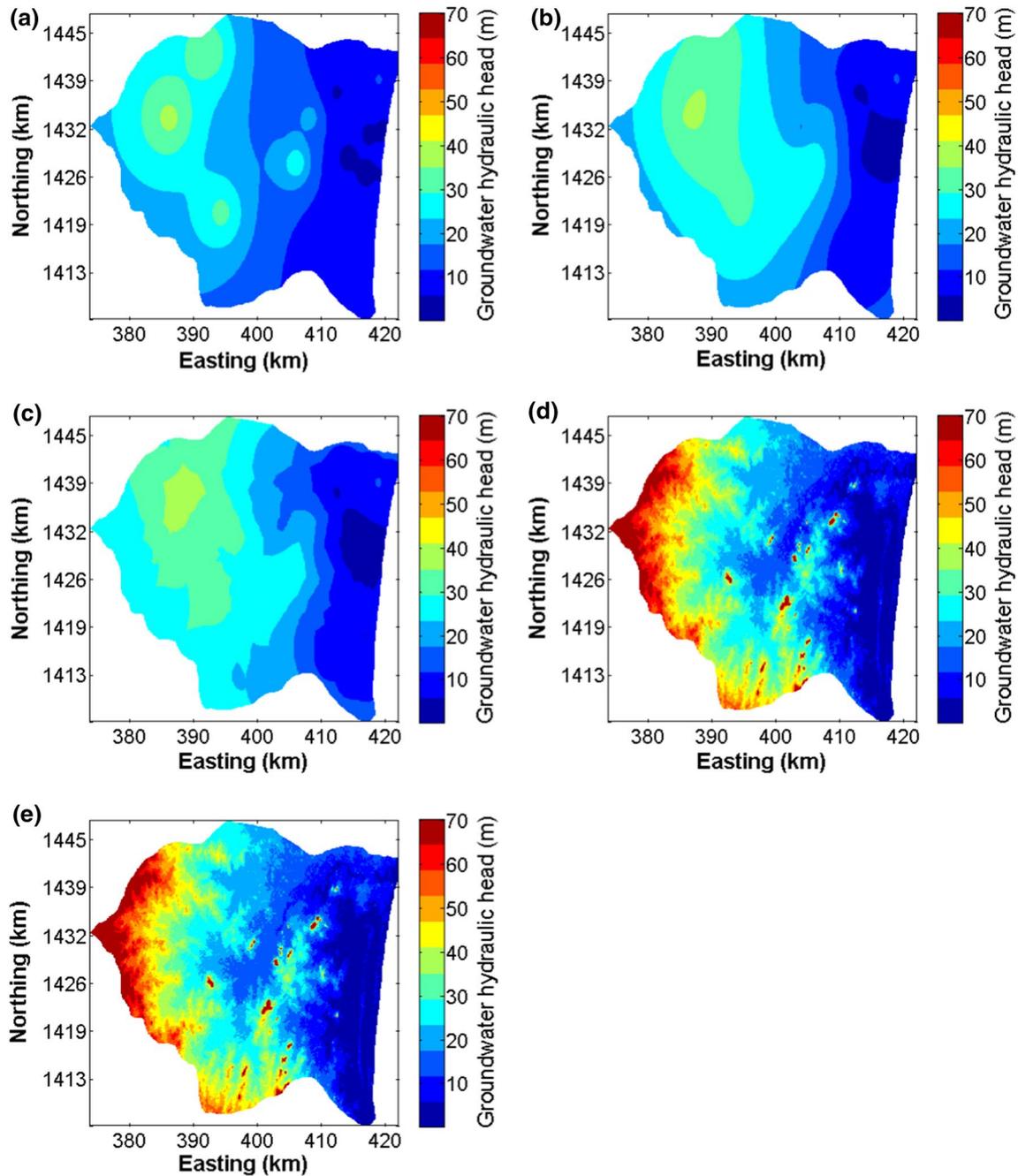
Fig. 11 Groundwater hydraulic head predictions by **a** IDW, **b** OK, **c** RK-DS, **d** RK-DEM, and **e** RK-MGWL using post-monsoon hydraulic head data

northeast and southwest monsoon seasons increase the groundwater table.

A significant reduction in the total prediction error variance (0 to 2.5 m^2) was observed for the RK-MGWL method over OK, RK-DS, and RK-DEM methods (Fig. 12). When compared to pre-monsoon prediction error, the post-monsoon prediction error for the RK-DEM method was comparatively higher which was ranged up to 14 m^2

against less than 6 m^2 in the pre-monsoon season analysis (Fig. 9c and Fig. 12c). However, a significant reduction in the prediction error variance for the RK-MGWL method in both pre-monsoon and post-monsoon season groundwater hydraulic head analysis (Figs. 9d and 12d) show that the proposed method outperforms over other traditional methods in all ranges of groundwater level values irrespective of the seasons.
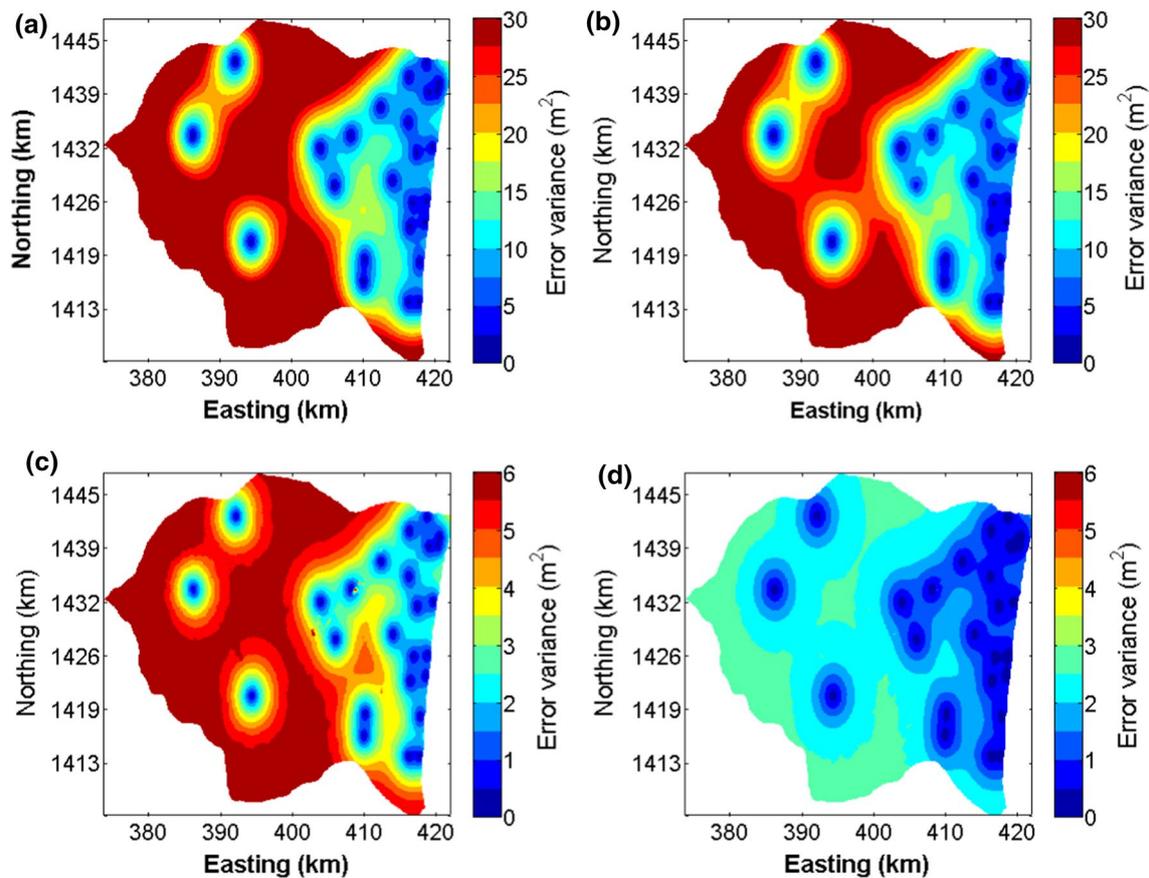
Fig. 12 Total error variance created by **a** OK, **b** RK-DS, **c** RK-DEM and **d** RK-MGWL models using post-monsoon hydraulic head data

Cross-validation statistics for the developed models

The overall model accuracy was calculated by LOOV analysis after iterating the respective kriging model predictions continuously after leaving one observation at each iteration and repeating the process for 29 repetitions for 29 well sites. Figure 13a, b shows the prediction and observation agreement plots for different interpolation methods, including IDW and other kriging methods. The effectiveness of the proposed RK method was observed as the predicted
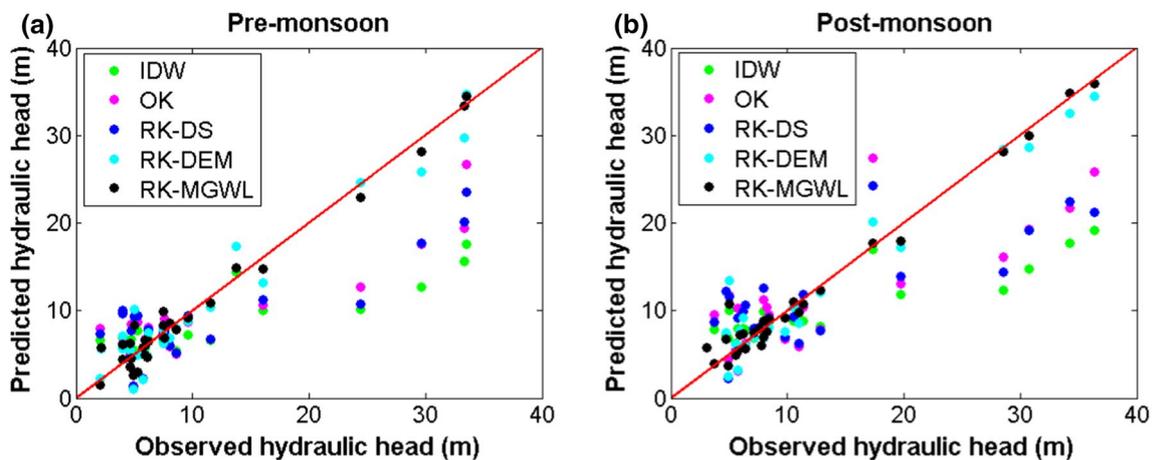
Fig. 13 Observed and predicted groundwater hydraulic head values by different interpolation methods using **a** pre-monsoon and **b** post-monsoon datasets of the year 2007

values were in close agreement with the observed values as compared to the other methods (Fig. 13). The IDW method severely under predicted the actual values at higher elevations, especially, in the western part of the basin. This could be because the IDW is based on the distance-based weighting function and very few points were available to interpolate most of the grid cells in the higher elevation regions from the western parts of the basin. In general, all the developed methods were performing better under relatively shallow groundwater table regions. At the same time, there was considerable uncertainty when the models predicted the groundwater levels at deeper groundwater zones of the basin except for the proposed RK-MGWL method (Fig. 13). The RK-MGWL method predicts the groundwater level surface with greater accuracy even at the higher elevated regions in the basin as the trend function used in the method effectively models the rising and lowering trends in the groundwater level data based on the season-specific trend model parameters.

The prediction results at 29 observation well locations and corresponding observed groundwater hydraulic head values were used in calculating the model performance statistical indices such as ME, MSE, MAE, RMSE, and R^2. The LOOV statistics ($n = 29$) for the respective kriging methods is given in Table 3. The results from Table 3 show that the proposed method (RK-MGWL) outperforms over other methods (IDW, OK, RK-DS, and RK-DEM) with respect to ME, MSE, MAE, RMSE, and R^2 indices values. The performance of the IDW method was observed poor among other methods as it severely underestimated the groundwater hydraulic heads both in pre-monsoon and post-monsoon seasons. The RK-DEM model performance was comparatively better than OK and RK-DS methods. However, it predicts the groundwater hydraulic heads with some considerable level of uncertainty when compared to the RK-MGWL method. For example, RK-DEM method RMSE values were calculated as 2.20 m and 2.51 m during pre-monsoon and post-monsoon seasons, respectively. These RMSE values were

almost 60% and 40% more as compared to the proposed RK-MGWL method (1.37 m and 1.75 m) during pre-monsoon and post-monsoon seasons, respectively (Table 3).

Conclusions

Spatial modeling of groundwater hydraulic heads in arid and semiarid regions is crucial for optimal management of groundwater resources, thus ensuring the sustainability of groundwater resources. Spatial modeling of groundwater levels is often attempted by several methods such as physical-based deterministic functions and stochastic-based kriging techniques. However, the traditional kriging interpolation methods have a considerable level of uncertainty associated with the predictions due to poor accountability of the trend component in the groundwater level datasets. An improved RK method with MGWL-based trend function has been proposed in the present study. The proposed trend function was formulated based on the MGWL variable, which was calculated based on long-term season-specific averaging of the groundwater levels. The effectiveness of the proposed kriging model was demonstrated against the traditional kriging and IDW methods in Adyar basin. The major conclusions of the present study are summarized below:

- The sill variance parameter was very high for OK and RK-DS methods as compared to RK-DEM and RK-MGWL methods. The OK and RK-DS methods did not effectively capture the spatial trend in the groundwater level data. The RK-DEM method modeled the trend surface based on a linear model with DEM as an auxiliary variable. The RK-DEM-based trend function effectively filtered the trend component in the groundwater level data as the shallow groundwater levels nearly align with the basin topography. Therefore, the remaining residual variance was relatively lesser for RK-DEM method as compared to RK-DS and OK methods. The residual var-

Table 3 Performance assessment of various spatial interpolation models by a cross-validation technique

Season	Model	Model performance indices ($n = 29$)				
		ME (m)	MSE (m^2)	MAE (m)	RMSE (m)	R^2
Pre-monsoon	IDW	2.27	42.26	3.96	6.5	0.45
	OK	0.87	28.94	3.83	5.34	0.63
	RK-SD	0.98	32.20	4.08	5.67	0.58
	RK-DEM	0.10	4.87	1.66	2.20	0.94
	RK-MGWL	−0.10	1.87	0.97	1.37	0.98
Post-monsoon	IDW	2.34	44.84	4.26	6.69	0.47
	OK	0.96	31.66	4.08	5.62	0.62
	RK-SD	1.06	36.68	4.46	6.05	0.57
	RK-DEM	0.52	6.63	1.85	2.57	0.92
	RK-MGWL	−0.08	3.06	1.15	1.75	0.96

iance generated from RK-MGWL method was significantly lesser as compared to all other methods developed in the present study. The MGWL variable was a function of long-term monthly averaged values which was used as a perfect representation for filtering the trend component effectively in the groundwater level signals, thus reduced the residual variance significantly as compared to other methods.

- Groundwater hydraulic head predictions by RK-DEM and RK-MGWL methods showed a realistic groundwater level variation with respect to the ground surface as RK-DEM and RK-MGWL methods used DEM- and MGWL-based trend functions to filter the trend component of the groundwater hydraulic heads data. However, a unique variation in the groundwater level predictions by the RK-DEM and RK-MGWL methods was also highlighted in the low-lying regions of the basin which shows the clear difference between the two models (RK-DEM and RK-MGWL). The IDW, OK, and RK-DS methods mostly underestimated the groundwater hydraulic head values in higher altitude regions; thus, these methods are not capable of modeling the groundwater level surface in complex hydrogeological conditions.

- Total error variance developed by different interpolation methods showed that the RK-MGWL method estimated the least error variance values in the spatial domain as compared to all other models. The total error variance at the well locations was relatively lesser, while the error variance increased more than 50 m^2 in the case of OK and RK-DS methods. On the other hand, RK-DEM and RK-MGWL methods predicted the error variance map with the maximum value of 6 m^2 and 1 m^2, respectively. Interestingly, the RK-MGWL method predicted the error variance map with less than 1 m^2 almost uniform over the entire spatial domain which shows the ability of the proposed method to predict the groundwater level values precisely even under a sparse distribution of the measurement locations.

- Cross-validation statistics during both pre-monsoon and post-monsoon groundwater level predictions show that the proposed kriging method outperforms over traditional kriging and IDW methods in terms of ME, MSE, MAE, RMSE, and R^2 values. For example, the calculated RMSE values for the proposed method were about 1.37 m and 1.75 m during pre-monsoon and post-monsoon seasons, respectively. But, the same index (RMSE) for RK-DEM and other methods was calculated nearly two to four times more than the RK-MGWL method.

The proposed method was applied to only one geographic region due to data limitations. The is a scope to evaluate the effectiveness of the proposed method in different geographical regions with different climatic and hydrogeologic

settings. Although the proposed method has been successfully validated in the spatial domain, it can also be used under spatiotemporal groundwater level modeling. In the spatiotemporal modeling, the proposed method can be used in the spatial domain for interpolating the groundwater levels in combination with the temporal prediction of groundwater levels by the time series models.

Data availability statement MATLAB scripts used for kriging and IDW models are available in the following GitHub repository: https://github.com/mohanasundaram1986/kriging

Compliance with ethical standards

Conflict of interest There is no conflict of interest among the authors of this paper.

References

Adhikary PP, Dash CJ (2017) Comparison of deterministic and stochastic methods to predict spatial variation of groundwater depth. Appl Water Sci 7:339–348. https://doi.org/10.1007/s13201-014-0249-8

Ahmadi SH, Sedghamiz A (2008) Application and evaluation of kriging and cokriging methods on groundwater depth mapping. Environ Monit Assess 138:357–368. https://doi.org/10.1007/s10661-007-9803-2

Amini MA, Torkan G, Eslamian S, Zareian MJ, Adamowski JF (2019) Analysis of deterministic and geostatistical interpolation techniques for mapping meteorological variables at large watershed scales. Acta Geophys 67:191–203. https://doi.org/10.1007/s11600-018-0226-y

Arun PV (2013) A comparative analysis of different DEM interpolation methods. Egypt J Remote Sens Space Sci 16:133–139. https://doi.org/10.1016/j.ejrs.2013.09.001

Bierkens MFP, Knotters M, van Geer FC (1999) Calibration of transfer function-noise models to sparsely or irregularly observed time series. Water Resour Res 35:1741–1750. https://doi.org/10.1029/1999WR900083

Bierkens MFP, Knotters M, Hoogland T (2001) Space-time modeling of water table depth using a regionalized time series model and the Kalman Filter. Water Resour Res 37:1277–1290. https://doi.org/10.1029/2000WR900353

Chung JW, Rogers JD (2012) Interpolations of groundwater table elevation in dissected uplands. Ground Water 50:598–607. https://doi.org/10.1111/j.1745-6584.2011.00889.x

Clark I (1979) Practical geostatistics. Applied Science Publishers Ltd, London

Cooper RM, Istok JD (1988) Geostatistics applied to groundwater contamination. I: Methodology. J Hydraul Eng 114:270–286

Dash JP, Sarangi A, Singh DK (2010) Spatial variability of groundwater depth and quality parameters in the National Capital Territory of Delhi. Environ Manag 45:640–650. https://doi.org/10.1007/s00267-010-9436-z

Daya AA, Bejari H (2015) A comparative study between simple kriging and ordinary kriging for estimating and modeling the Cu concentration in Chehlkureh deposit, SE Iran 6003–6020. Doi:10.1007/s12517-014-1618-1

Delbari M, Motlagh MB, Amiri M (2013) Spatio-temporal variability of groundwater depth in the Eghlid aquifer in southern. Iran 17:105–114

Desbarats AJ, Logan CE, Hinton MJ, Sharpe DR (2002) On the kriging of water table elevations using collateral information from a digital elevation model. J. Hydrol. 255:25–38. https://doi.org/10.1016/S0022-1694(01)00504-2

Gong G, Mattevada S, O'Bryant SE (2014) Comparison of the accuracy of Kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. Environ Res 130:59–69. https://doi.org/10.1016/j.envres.2013.12.005

Gundogdu KS, Guney I (2007) Spatial analyses of groundwater levels using universal kriging. J Earth Syst Sci. https://doi.org/10.1007/s12040-007-0006-6

Hengl T, Heuvelink GBM, Stein A (2003) Comparison of Kriging with external drift and regression-Kriging. Technical note, ITC, Available on-line at https://www.itc.nl/library/Academicoutput/

Isaaks EH, Srivastava RM (1989) Applied geostatistics. Oxford University Press, New York

Kumar V (2007) Optimal contour mapping of groundwater levels using universal kriging: a case study. Hydrol Sci J 52:1038–1050. https://doi.org/10.1623/hysj.52.5.1038

Kumar D, Ahmed S (2003) Seasonal behaviour of spatial variability of groundwater level in a granitic aquifer in monsoon climate. Curr Sci 84(2): 188–196. Retrieved February 6, 2020, from www.jstor.org/stable/24108097

Kumar V, Remadevi (2006) Kriging of groundwater levels—a case study. J Spatial Hydrol 6:81–94

Landrum C, Castrignanó A, Zourarakis D, Mueller T (2016) Assessing the time stability of soil moisture patterns using statistical and geostatistical approaches. Agric Water Manag 177:118–127. https://doi.org/10.1016/j.agwat.2016.07.013

Ma T-S, Sophocleous M, Yu Y-S (1999) Geostatistical applications in ground-water modeling in south-central kansas. J. Hydrol. Eng 4(1):57

Marache, A., Lastennet, R., Breysse, D., (2016) Geostatistical investigations for suitable mapping of the water table: the Bordeaux case (France). Hydrogeol J 24:231–248. https://doi.org/10.1007/s10040-015-1316-4

Martínez-Murillo JF, Hueso-González P, Ruiz-Sinoga JD (2017) Topsoil moisture mapping using geostatistical techniques under different Mediterranean climatic conditions. Sci Total Environ 595:400–412. https://doi.org/10.1016/j.scitotenv.2017.03.291

Mhamad AJ (2019) Using regression Kriging to analyze groundwater according to depth and capacity of wells. UHD J. Sci. Technol. 3(1):39–47. https://doi.org/10.21928/uhdjst.v3n1y2019

Mohanasundaram S, Narasimhan B, Suresh Kumar G (2017) Transfer function noise modelling of groundwater level fluctuation using threshold rainfall-based binary-weighted parameter estimation approach. Hydrol Sci J 62(1):36–49. https://doi.org/10.1080/02626667.2016.1171325

Mohanasundaram S, Suresh Kumar G, Narasimhan B (2019) A novel deseasonalized time series model with an improved seasonal estimate for groundwater level predictions. H2Open J. 2:25–44. https://doi.org/10.2166/h2oj.2019.022

Nikroo L, Kompani-Zare M, Sepaskhah AR, Fallah Shamsi SR (2010) Groundwater depth and elevation interpolation by kriging methods in Mohr Basin of Fars province in Iran. Environ Monit Assess. https://doi.org/10.1007/s10661-009-1010-x

Plouffe CCF, Robertson C, Chandrapala L (2015) Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: a case study of Sri Lanka. Environ Model Softw 67:57–71. https://doi.org/10.1016/j.envsoft.2015.01.011

Rivest M, Marcotte D, Pasquier P (2008) Hydraulic head field estimation using kriging with an external drift: a way to consider conceptual model information. J Hydrol 361:349–361. https://doi.org/10.1016/j.jhydrol.2008.08.006

Rizo-Decelis LD, Pardo-Igúzquiza E, Andreo B (2017) Spatial prediction of water quality variables along a main river channel, in presence of pollution hotspots. Sci Total Environ 605–606:276–290. https://doi.org/10.1016/j.scitotenv.2017.06.145

Shirmohammadi B, Vafakhah M, Moosavi V, Moghaddamnia A (2013) Application of several data-driven techniques for predicting groundwater level. Water Resour Manag 27:419–432. https://doi.org/10.1007/s11269-012-0194-y

Shtiliyanova A, Bellocchi G, Borras D, Eza U, Martin R, Carrère P (2017) Kriging-based approach to predict missing air temperature data. Comput Electron Agric 142:440–449. https://doi.org/10.1016/j.compag.2017.09.033

Sun Y, Kang S, Li F, Zhang L (2009) Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China. Environ Model Softw 24:1163–1170. https://doi.org/10.1016/j.envsoft.2009.03.009

Tankersley CD, Graham WD, Hatfield K (1993) Comparison of univariate and transfer function models of groundwater fluctuations. Water Resour Res 29:3517–3533. https://doi.org/10.1029/93WR01527

Theodoridou PG, Varouchakis EA, Karatzas GP (2017) Spatial analysis of groundwater levels using fuzzy logic and geostatistical tools. J Hydrol 555:242–252. https://doi.org/10.1016/j.jhydrol.2017.10.027

Theodossiou N, Latinopoulos P (2006) Evaluation and optimisation of groundwater observation networks using the Kriging methodology. Environ Model Softw 21:991–1000. https://doi.org/10.1016/j.envsoft.2005.05.001

Varouchakis EA, Corzo GA, Karatzas GP, Kotsopoulou A (2018) Spatio-temporal analysis of annual rainfall in Crete. Greece Acta Geophys 66:319–328. https://doi.org/10.1007/s11600-018-0128-z

Varouchakis EA, Hristopulos DT (2013) Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables. Adv Water Resour 52:34–49. https://doi.org/10.1016/j.advwatres.2012.08.002

Varouchakis EA, Hristopulos DT, Karatzas GP (2012) Improving Kriging of groundwater level data using nonlinear normalizing transformations: a field application. Hydrol Sci J 57(7):1404–1419. https://doi.org/10.1080/02626667.2012.717174

Vereecken H, Huisman JA, Pachepsky Y, Montzka C, van der Kruk J, Bogena H, Weihermüller L, Herbst M, Martinez G, Vanderborght J (2014) On the spatio-temporal dynamics of soil moisture at the field scale. J. Hydrol. 516:76–96. https://doi.org/10.1016/j.jhydrol.2013.11.061

Wagner PD, Fiener P, Wilken F, Kumar S, Schneider K (2012) Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. J Hydrol 464–465:388–400. https://doi.org/10.1016/j.jhydrol.2012.07.026

WRO (Water Resources Organization) (2007) Micro level study Chennai basin. Volume-I Institute of Water Studies, Taramani, Chennai

Wu T, Li Y (2013) Spatial interpolation of temperature in the United States using residual Kriging. Appl Geogr 44:112–120. https://doi.org/10.1016/j.apgeog.2013.07.012

Yin S, Xiao Y, Gu X, Hao Q, Liu H, Hao Z, Meng G, Pan X, Pei Q (2019) Geostatistical analysis of hydrochemical variations and nitrate pollution causes of groundwater in an alluvial fan plain. Acta Geophys 67:1191–1203. https://doi.org/10.1007/s11600-019-00302-5

Zhu K, Cui Z, Jiang B, Yang G, Chen Z, Meng Q, Yao Y (2013) A DEM-based residual Kriging model for estimating groundwater levels within a large-scale domain: a study for the Fuyang River Basin. Clean Technol Environ Policy. https://doi.org/10.1007/s10098-012-0563-5