



# Geostatistical analysis of groundwater levels in a mining area with three active mines

A. Pavlides<sup>1</sup> · E. A. Varouchakis<sup>1</sup> · D. T. Hristopulos<sup>2</sup>

Received: 8 November 2022 / Accepted: 30 June 2023 / Published online: 25 August 2023  
© The Author(s) 2023

## Abstract

Mining activities can significantly impact groundwater reservoirs in their vicinity. Different approaches have been employed, with varying success, to investigate the spatial variability of groundwater levels in mining areas. Typical problems include the small sample size, the non-Gaussian distribution of the data, and the clustering of sample locations near the mines. These conditions complicate the estimation of spatial dependence. Under sparse and irregular sampling conditions, stochastic methods, which can provide estimates of prediction uncertainty, are preferable to deterministic ones. This research focuses on the comparison of two stochastic methods, stochastic local interactions (SLI) and universal Kriging (UK), using water level data from 72 locations around three mines in Northern Greece. UK is a well-known, variogram-based geostatistical method, while SLI is a computationally efficient kernel-based method that can cope with large spatial datasets. The non-Gaussian distribution of the data is handled by means of a flexible, data-driven Gaussian anamorphosis method that uses kernel functions. The spatial prediction performance of both methods is assessed based on cross-validation. UK performs better than SLI, due to the fact that the former incorporates a linear trend function. On the other hand, a comparison of the two methods using data from a single mine that contains only 28 measurement locations shows that SLI performs slightly better than UK. The prediction uncertainties for both methods are also estimated and compared. The results suggest that SLI can provide better estimates than classical geostatistical methods for small sample sizes that do not allow reliable estimation of the variogram model.

**Keywords** Geostatistics · Groundwater monitoring · Non-Gaussian · Stochastic local interaction · Gaussian anamorphosis

## Introduction

The quantity, quality, and the degree of spatiotemporal variability of groundwater are key factors for the wealth of human communities and ecosystems; therefore, they play a significant role in effective groundwater resources management. Both human interventions, such as excessive pumping, and natural processes, such as a decline in recharge, can cause fluctuations in groundwater levels. Level variations could be long-term (on the order of years) or short-term

(on the order of months). Global studies confirm the connection between excessive groundwater pumping and variations in groundwater levels (Islam et al. 2022; Tatas et al. 2022). Dewatering efforts for underground mining operations significantly reduce groundwater levels for a long time and in many cases may dry up nearby springs. Near-mine ecosystems typically depend on groundwater for survival. Water pumping for mining activities can lead to groundwater level drops which can be harmful to local communities and ecosystems (Schrader and Winde 2015; Davies et al. 2020). Therefore, monitoring the spatial distribution of groundwater levels around mining areas is an important goal for management purposes.

Various geostatistical methods have been used to investigate changes in the spatial distribution of groundwater levels. Some studies employ Kriging methods using only primary information (Budiman et al. 2022; Evans et al. 2020), while others involve auxiliary variables (Varouchakis and Hristopulos 2013a, b; Zirakbash et al. 2020;

Published in the special issue “Geostatistics and hydrogeology”.

✉ E. A. Varouchakis  
evarouchakis@tuc.gr

<sup>1</sup> School of Mineral Resources Engineering, Technical University of Crete, Chania, Crete 73100, Greece

<sup>2</sup> School of Electrical and Computer Engineering, Technical University of Crete, Chania, Crete 73100, Greece

Calzolari and Ungaro 2012; An et al. 2015) or implement hybrid machine learning approaches (Parasyris et al. 2021; Manzione and Castrignano 2019; Theodoridou et al. 2017; Tapoglou et al. 2014). In mining areas, groundwater levels distribution is usually investigated through the use of Kriging methods such as ordinary Kriging, indicator Kriging and co-Kriging (Dai et al. 2011; Abiye et al. 2018; Keegan-Treloar et al. 2021).

Geostatistical methods typically rely on the existence of data that sufficiently sample the spatial variability in the area of interest. In mining areas, the monitoring stations are often clustered around small areas of interest (pits, mines or mine sectors). This practice prevails because such areas require careful monitoring due to increased environmental risks. The clustered pattern of the measurement locations hinders the reliable estimation of the variogram model, especially if the number of data points inside each cluster is small (Goovaerts 1997; Kovitz and Christakos 2004). Another problem commonly faced by geostatistical methods is the skewness of the data distribution. This is usually handled by means of nonlinear transforms (e.g., lognormal, Box-Cox) which aim to normalize the marginal data distribution (this procedure is known as Gaussian anamorphosis; Wackernagel 2003; Chilès and Delfiner 2012).

This work has two goals: First, it presents an alternative approach for spatial groundwater levels analysis that can overcome the difficulties related to estimating spatial dependence. This approach is based on the *stochastic local interaction* (SLI) model. This model imposes spatial correlations by means of local couplings (interactions); the SLI model can increase the computational efficiency of spatial prediction for correlated data (Hristopulos 2015; Hristopulos et al. 2021). Secondly, a recently proposed data-driven (nonparametric), kernel-based approach is used to conduct Gaussian anamorphosis (Pavlidis et al. 2022; Agou et al. 2022). This method is more flexible than nonlinear transforms based on explicit functions and does not suffer from the problems that Hermite polynomials face in the tails of the distribution.

Groundwater level data (meters below surface) from three mines operating in the same region (northern Greece), are studied. Groundwater is used to support mining activity in the study area. Groundwater levels are investigated by means of universal Kriging (UK) and SLI. Uncertainty estimates for both SLI and UK are provided. Prediction intervals (at 5 and 95% probability levels) are estimated. Leave-one-out cross-validation (LOOCV) measures indicate that the performance of UK and SLI is adequate considering the small sample size—although both methods slightly overestimate a few low values. UK performs slightly better than SLI if the entire dataset of the three mines is taken into account; however, SLI performs better than UK even for smaller sample sizes (individual mines). MATLAB code for SLI is freely available at Hristopulos (2020b).

The rest of this paper is structured as follows: section “**Theory and methods**” reviews the theory for the spatial prediction methods employed (UK and SLI) and for Gaussian anamorphosis; section “**Data description and preprocessing**” includes the data description, exploratory analysis and data normalization; section “**Results: universal Kriging**” presents the application of UK to the groundwater level data including cross-validation analysis; section “**Results: SLI**” focuses on the application of SLI over the entire area and separately to one of the three mines, along with the respective cross-validation measures. Lastly, section “**Discussion**” comments on the results, and the final section presents conclusions and open questions for further study.

## Theory and methods

The spatial distribution of variables is used using the mathematical framework of random fields (Adler 1981; Christakos 1992; Chilès and Delfiner 2012; Hristopulos 2020a). A random field is a collection of dependent random variables distributed over the spatial domain of interest. Herein, the random field concept is used to model groundwater levels.

A scalar random field  $X(\mathbf{s})$  where  $\mathbf{s} \in \mathbb{R}^d$  is the position vector, represents a random function  $\mathcal{D} \subset \mathbb{R}^d \rightarrow \mathbb{R}$ , where  $\mathcal{D}$  is the spatial domain; in this study  $d = 2$  and  $\mathbf{s} = (s_1, s_2)^\top$ , where  $\top$  denotes the transpose of a vector. A random field is fully determined by means of the  $n$ -point joint probability density functions, where  $n = 1, 2, \dots$  denotes a positive integer (the set of all positive integers is denoted by  $\mathbb{N}$ ). The Gaussian assumption is often used for the joint density functions.

The data are assumed to represent a partial sample of the random field; they are denoted by the set of sampling sites  $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$  and the respective water level measurements  $\{x_1, \dots, x_N\}$  (lowercase letters denote specific realizations of the random field). The field is reconstructed on a rectangular map grid  $\mathbb{G}$  which comprises the set of points (nodes)  $\{\mathbf{z}_1, \dots, \mathbf{z}_P\}$ . Estimation points will be denoted in general by  $\mathbf{u} \in \mathbb{R}^d$ , where  $\mathbf{u}$  refers either to a point in  $\mathbb{G}$  or one of the sampling points (e.g., for cross-validation analysis).

For practical purposes, random fields are often decomposed into a deterministic part (trend) and a stochastic part (fluctuation) as follows:

$$X(\mathbf{s}) = m_x(\mathbf{s}) + X'(\mathbf{s}) \quad (1)$$

The function  $m_x(\mathbf{s})$  is assumed to represent the mathematical expectation of the random field, i.e.,  $m_x(\mathbf{s}) = \mathbb{E}[X(\mathbf{s})]$ . The expectation  $\mathbb{E}[X(\mathbf{s})]$  of a random field at the point  $\mathbf{s}$  represents a stochastic average over all probable configurations of the field (as defined by means of the joint probability distribution; Hristopulos 2020a). Hence,  $m_x(\mathbf{s})$  is a deterministic function and typically it reflects slow spatial variations

(trend) of the random field  $X(\mathbf{s})$ ; usually,  $m_x(\mathbf{s})$  is modeled as a polynomial. Herein, only first-degree polynomial trends are considered, i.e.,

$$m_x(\mathbf{s}) = a_0 + a_1s_1 + a_2s_2 \tag{2}$$

where  $\mathbf{A} = (a_0, a_1, a_2)^\top$  is the vector of trend coefficients, and  $\mathbf{s} = (s_1, s_2)$  are the coordinates of the two-dimensional (2D) position vector  $\mathbf{s}$ .

The fluctuation  $X'(\mathbf{s})$  is a random field that represents the finer-scale, stochastic component of  $X(\mathbf{s})$ . It is obtained by removing the trend from  $X(\mathbf{s})$  and is thus equal to the field's residuals (Pavlidis et al. 2015). Since it is assumed that  $m_x(\mathbf{s}) = \mathbb{E}[X(\mathbf{s})]$ , it follows that  $\mathbb{E}[X'(\mathbf{s})] = 0$ .

A random field is considered *stationary* if the joint  $n$ -point probability distributions (for all  $n \in \mathbb{N}$ ) do not depend on the spatial location but only the relative configuration of the points. Since this condition is not easily testable, the condition of weak stationarity is used in practice. Weak stationarity means that the expectation of the random field is constant and the covariance is a function purely of the time lag (not the particular instant in time). A stationary Gaussian random field is thus fully determined from the mean function  $m_x(\mathbf{s})$  and the covariance function  $c_x(\mathbf{r})$ , where  $\mathbf{r} \in \mathbb{R}^d$  is the space lag between two points  $\mathbf{s}$  and  $\mathbf{s}' = \mathbf{s} + \mathbf{r}$  and  $c_x(\cdot)$  is a positive-definite function.

It is assumed that the data sites comprise the set of points  $\{\mathbf{s}_i\}_{i=1}^n$  while the groundwater level measurements at these sites will be denoted by the data vector  $\mathbf{x} = (x_1, \dots, x_n)^\top$ .

### Variogram modeling

The models attempting to represent spatial processes rely on the spatial correlations between the data. These correlations are reflected in the covariance function that characterizes the random field; however, for practical reasons, it is more convenient to estimate the variogram function instead of the covariance function (Chilès and Delfiner 2012; Kitanidis and Vomvoris 1983).

The variogram function  $\gamma_x(\mathbf{r})$  for stationary random fields is connected to the covariance function  $c_x(\mathbf{r})$  as shown in Eq. (3).

$$\gamma_x(\mathbf{r}) = c_x(0) - c_x(\mathbf{r}) \tag{3}$$

where  $\mathbf{r}$  is the distance vector between two points. Thus, the variogram function can be used interchangeably with the covariance function if the field of the fluctuations  $X'$  is considered a stationary field.

The variogram function is defined as follows

$$\gamma_x(\mathbf{r}) = \frac{1}{2} \text{Var}[X(\mathbf{s}) - X(\mathbf{s} + \mathbf{r})] \tag{4}$$

where  $\text{Var}$  is the variance operator, i.e.,  $\text{Var}(X) = \mathbb{E}[X^2(\mathbf{s})] - \mathbb{E}^2[X(\mathbf{s})]$  Olea (2006).

The empirical (experimental) variogram is obtained from the data using Matheron's moment estimator (Matheron 1962) or Cressie's robust variogram estimator (Cressie 1993). The empirical variogram is then fitted to a theoretical model in order to account for all possible pair distances. Various admissible variogram models can be found in Goovaerts (1997); Chilès and Delfiner (2012); Hristopulos (2020a). The tested variogram models are fitted to the empirical variogram, using the weighted least squares method (Olea 2006). The spherical variogram model which is used herein satisfies the equation

$$\gamma_x(\mathbf{r}) = \begin{cases} \sigma^2 [1.5(|\mathbf{r}|/\xi) - 0.5(|\mathbf{r}|/\xi)^3], & |\mathbf{r}| \leq \xi \\ \sigma^2, & |\mathbf{r}| \geq \xi \end{cases} \tag{5}$$

where  $\sigma^2$  is the variance of the random field,  $|\mathbf{r}|$  is the spatial lag, and  $\xi$  is the correlation length under the assumption of statistical isotropy (i.e., lack of directional preference; Hristopulos 2020a).

### Kriging

Kriging refers to a group of stochastic spatial interpolation methods. They are by construction linear, unbiased, and minimum variance estimators. Kriging methods have a large range of applications in the mining and environmental sectors (Cressie 1990; Varouchakis et al. 2018; Varouchakis and Hristopulos 2013b). Kriging provides significant benefits compared to deterministic interpolation methods (Gong et al. 2014; Varouchakis and Hristopulos 2013a; Pavlidis et al. 2015).

The value of the random field  $X(\mathbf{u})$  at an unmeasured location  $\mathbf{u} \in \mathbb{G}$  is estimated by means of a linear combination of the measurements at  $n(\mathbf{u})$  nearby points  $\mathbf{s}_1(\mathbf{u}), \dots, \mathbf{s}_{n(\mathbf{u})}$ , where  $\mathbf{s}_{i(\mathbf{u})} \in \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$  is a neighbor of  $\mathbf{u}$  for all  $i = 1, \dots, n(\mathbf{u})$ . A map of the spatial distribution of the field is obtained by repeating the estimation process at every node of a suitably selected grid. Such maps can be accompanied by estimates of the prediction variance at each point. The variance can adequately represent the prediction uncertainty if the data probability distribution is Gaussian. If the data follow a skewed probability distribution, the Kriging-based uncertainty estimates are not reliable (Agou 2016).

Simple Kriging (SK) assumes that the random field is stationary and its mean,  $m_x$ , is known. The SK estimator is obtained by means of the following equation (Krige 1951; Cressie 1990):

$$\hat{x}(\mathbf{u}) = m_x + \sum_{i=1}^{n(\mathbf{u})} \lambda_i(\mathbf{u}) (x_i - m_x) \tag{6}$$

In Eq. (6),  $\lambda_i(\mathbf{u})$  are the Kriging weights at the target point  $\mathbf{u}$ . The Kriging weights are calculated by solving the following linear system

$$\mathbf{C}_x \mathbf{\Lambda} = \mathbf{C}_u \tag{7}$$

where  $\mathbf{C}_x$  is the covariance (or variogram) matrix of the fluctuation random field  $X'(\cdot)$  at the data locations, i.e.,  $[\mathbf{C}_x]_{i,j} = c_x(\mathbf{s}_i - \mathbf{s}_j)$  for  $i, j = 1, \dots, n(\mathbf{u})$ ,  $\mathbf{\Lambda} = (\lambda_1(\mathbf{u}), \dots, \lambda_n(\mathbf{u}))^T$  is the vector of the Kriging weights, and  $\mathbf{C}_u = (c_x(\mathbf{u} - \mathbf{s}_1), \dots, c_x(\mathbf{u} - \mathbf{s}_{n(\mathbf{u})}))^T$  is the covariance (or variogram) vector between the unknown point  $\mathbf{u}$  and each of the  $n(\mathbf{u})$  neighbor points that contribute to the estimate (Chilès and Delfiner 2012). The Kriging estimate  $\hat{x}(\mathbf{u})$  at  $\mathbf{u}$  is obtained by replacing in Eq. (6) the  $x_i$  with the respective data values,  $m_x$  with the estimate of the mean, and the Kriging weights with the solution of Eq. (7) for  $\mathbf{\Lambda}$ .

### Universal Kriging

For many datasets the constant mean assumption is not suitable; even if the assumption is reasonable, it is not always possible to accurately determine the value of the mean (e.g., if the sample is small). Several variations of Kriging have been suggested to address this issue, such as ordinary, regression, and universal Kriging (Goovaerts 1997; Schabenberger and Gotway 2004; Cressie 1990).

In this study, universal Kriging (UK) is employed, which is also known as Kriging with a drift. This method assumes that the trend  $m_x(\mathbf{s})$  depends on the coordinates  $\mathbf{s}$  in terms of simple basis functions (Mesić Kiš 2016; Kitanidis 1997). Assuming a simple linear polynomial trend,  $m_x(\mathbf{s}) = a_0 + a_1 s_1 + a_2 s_2$ , where the  $a_0, a_1, a_2$  are real-valued, unknown coefficients, the UK estimator at location  $\mathbf{u} = (u_i, u_j)$  is

$$\hat{x}_{UK}(\mathbf{u}) = \sum_{k=1}^{n(\mathbf{u})} \lambda_k(\mathbf{u}) x(\mathbf{s}_k) \tag{8}$$

In Eq. (8), the polynomial coefficients of the trend  $m_x$  are considered unknown and should be estimated along with the covariance coefficients from the data, thus increasing the uncertainty of the prediction. The linear weights are calculated as the solution of Eq. (9)

$$\mathbf{C}_x \mathbf{B} = \mathbf{C}_u \tag{9}$$

where the vector  $\mathbf{B}$  contains both the  $n(\mathbf{u})$  weights and three Lagrange multipliers (one for each trend coefficient) used to enforce the no-bias condition. The covariance matrices  $\mathbf{C}_x$  and  $\mathbf{C}_u$  are expressed as follows

$$\begin{bmatrix} \sigma_x^2 & \dots & c_x(\mathbf{s}_1, \mathbf{s}_n) & 1 & s_{1:1} & s_{1:2} \\ c_x(\mathbf{s}_2, \mathbf{s}_1) & \dots & c_x(\mathbf{s}_2, \mathbf{s}_n) & 1 & s_{2:1} & s_{2:2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_x(\mathbf{s}_n, \mathbf{s}_1) & \dots & \sigma_x^2 & 1 & s_{n:1} & s_{n:2} \\ 1 & \dots & 1 & 0 & 0 & 0 \\ s_{1:1} & \dots & s_{n:1} & 0 & 0 & 0 \\ s_{1:2} & \dots & s_{n:2} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu_0 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} c_x(\mathbf{s}_1, \mathbf{u}) \\ c_x(\mathbf{s}_2, \mathbf{u}) \\ \vdots \\ c_x(\mathbf{s}_n, \mathbf{u}) \\ 1 \\ u_1 \\ u_2 \end{bmatrix} \tag{10}$$

where  $s_{k:1}$  and  $s_{k:2}$  are the coordinates of the point  $\mathbf{s}_k$ ,  $k = 1, \dots, n(\mathbf{u})$ . For convenience, the  $\mathbf{u}$  dependence of  $\lambda_k(\mathbf{u})$ ,  $k = 1, \dots, n(\mathbf{u})$ ,  $\mu_1(\mathbf{u})$ ,  $\mu_2(\mathbf{u})$  and  $n(\mathbf{u})$  is suppressed.

The covariance (or variogram) used in Eq. (10) should be based on the residuals which are obtained by subtracting the trend from the data). However, the trend is considered unknown, and this creates a cyclical problem. Methods to address this issue are suggested in Kitanidis (1997); Goovaerts (1997). If the trend is first estimated based on the data and then removed to obtain the residuals, a more appropriate methodology is regression Kriging.

In this study, the variogram model parameters are estimated by means of *maximum likelihood estimation* (MLE) Fletcher (2000); Hristopulos (2020a). The spherical variogram model, given by Eq. (5), is used. The maximization of the likelihood is equivalent to the minimization of the *negative logarithmic likelihood* (NLL). The latter is given by

$$NLL = \frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{C}_x^{-1}(\mathbf{x} - \mathbf{m}) + \frac{1}{2} \log(\det \mathbf{C}_x) + \frac{N}{2} \log(2\pi) \tag{11}$$

where  $\mathbf{x} = (x_1, \dots, x_N)^T$  is the data vector,  $\hat{\mathbf{m}}$  is the vector of the trend estimates at the sampling locations, i.e.,  $m_i = a_0 + a_1 s_i + a_2 s_j$ , and  $\det \mathbf{C}_x$  is the determinant of the covariance matrix  $\mathbf{C}_x$ . The trend parameters ( $a_0, a_1, a_2$ ) are estimated along with the parameters ( $\sigma^2, \xi$ ) of the covariance model by minimizing the NLL.

The minimum error variance of the UK prediction,  $\sigma_{UK}^2(\mathbf{u})$ , is obtained from the following equation (Kitanidis 1997, p. 127):

$$\sigma_{UK}^2(\mathbf{u}) = c_x(0) - \sum_{k=1}^{n(\mathbf{u})} \lambda_k(\mathbf{u}) c_x(\mathbf{u} - \mathbf{s}_k) - \mu_0(\mathbf{u}) - \mu_1(\mathbf{u})u_1 - \mu_2(\mathbf{u})u_2 \tag{12}$$

### Kernel functions

Kernel functions are routinely used in nonparametric estimation methods (Ghosh 2018). Herein, kernel functions are used for two purposes: (1) To define the SLI weights (see section “[Stochastic local interaction model](#)”), and (2) to estimate the *cumulative distribution function* (CDF) of the data. The SLI model is discussed in section “[Stochastic local interaction model](#)”. The CDF estimation is described in the following.

A kernel function  $K(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  satisfies the following requirements (Ghosh 2018):

- $K(x)$  is nonnegative, i.e.,  $K(x) \geq 0$  and symmetric function, i.e.,  $K(-x) = K(x)$  for all  $x \in \mathbb{R}$ .
- $K(x)$  is normalized so that  $\int_{-\infty}^{+\infty} K(x) dx = 1$ .
- If  $K(x)$  is a kernel function, the function  $K_\lambda(x) = \lambda K(\lambda x)$  is also a kernel  $\forall \lambda > 0$ .

The parameter  $\lambda$  adjusts the kernel’s range via the bandwidth parameter  $b = 1/\lambda$ . Larger values of  $b$  imply a longer kernel range. Table 1 summarizes some kernel functions that are used in this study. At this point, a comment on notation is necessary: the symbol  $h$  is typically used to denote the bandwidth. Since in the current study kernel functions are used for CDF estimation and in SLI models, to distinguish between the two kernel bandwidths  $b$  is used for CDF estimation ( $b$  has units of [X]) and  $h$  in SLI modeling ( $h$  has units of distance).

Next, the nonparametric CDF estimation procedure is described. Let  $\{x_{[i]}\}_{i=1}^N$  represent the ordered sample of the water level values  $\mathbf{x}$ : this is defined so that for any  $i \in \{1, \dots, N\}$  there is  $j \in \{1, \dots, N\}$  such that  $x_{[i]} = x_j$  and  $x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[N]}$ . Then, the nonparametric, kernel-based estimate of the CDF (KCDE),  $\hat{F}_K(\cdot)$ , can be obtained from the following weighted sum (Pavlidis et al. 2022)

$$\hat{F}_K(x) = \sum_{i=1}^N \frac{1}{N} \tilde{K}\left(\frac{x - x_{[i]}}{b}\right) \tag{13}$$

where  $\tilde{K}(\cdot)$  is the CDF kernel step defined by means of the integral (Ghosh 2018; Hristopulos 2020a)

$$\tilde{K}\left(\frac{x - x_{[i]}}{b}\right) = \frac{1}{b} \int_{-\infty}^x K\left(\frac{x' - x_{[i]}}{b}\right) dx' \tag{14}$$

**Table 1** Common kernel functions

| Kernel       | $K(x)$  |
|--------------|---|
| Uniform      | $\frac{x}{2} \mathbb{I}( x  \leq 1)$                |
| Triangular   | $(1 -  x ) \mathbb{I}( x  \leq 1)$                  |
| Epanechnikov | $\frac{3}{4}(1 - x^2) \mathbb{I}( x  \leq 1)$       |
| Biweight     | $\frac{15}{16}(1 - x^2)^2 \mathbb{I}( x  \leq 1)$   |
| Spherical    | $(1 - 1.5 x  + 0.5 x ^3) \mathbb{I}( x  \leq 1)$    |
| Triweight    | $(1 - x^2)^3 \mathbb{I}( x  \leq 1)$                |
| Tricubic     | $\frac{70}{81}(1 -  x ^3)^3 \mathbb{I}( x  \leq 1)$ |

$\mathbb{I}(|x| \leq 1)$  is an indicator function used to define the range of compactly supported kernels:  $\mathbb{I}(|x| \leq 1) = 1$  when  $|x| \leq 1$  and  $\mathbb{I}(|x| \leq 1) = 0$ , if  $x > 0$

The KCDE-based  $\hat{F}_K(x)$  has several advantages compared to the empirical staircase CDF estimate  $\hat{F}(x)$ . For example, while  $\hat{F}(x)$  is discontinuous,  $\hat{F}_K(x)$  provides smoothed kernel steps by adapting the bandwidth  $b$  to the data. A continuous estimate of the CDF is especially useful for numerical simulations (Pavlidis et al. 2022). For the KCDE, Eq. (13), a plug-in bandwidth  $b$  is used, based on the algorithm developed by Botev et al. (2010). This bandwidth selection method was successfully tested for non-Gaussian data (Pavlidis et al. 2022).

### Stochastic local interaction model

The SLI model employs a local representation that improves the computational efficiency of spatial prediction for large datasets (Hristopulos 2015, 2020a; Hristopulos et al. 2021). SLI is based on a conditional probability density function (PDF) defined by means of an energy functional that involves local interactions between neighboring sites. The energy functional represents the “probability cost” for specific spatial configurations (patterns) of field values. The local interactions are implemented by means of kernel functions with locally adaptive kernel bandwidths  $h$ .

The SLI model is expressed mathematically by means of a precision matrix  $\mathbf{J}$ . The term “precision matrix” refers to the inverse of the covariance matrix. In classical geostatistical analysis, the precision matrix is obtained by inverting the covariance. In contrast, in SLI the precision matrix is constructed first. This representation leads to semianalytical expression for the prediction (conditional mean) of the random field, while the conditional variance is easily expressed in terms of precision matrix elements. Assuming a Gaussian distribution, the conditional mean and variance fully determine the conditional probability distribution at the prediction point. Hence, SLI prediction avoids the computationally costly inversion of the covariance matrix required by Kriging methods.

The PDF of an SLI model is given by means of the following Boltzmann-Gibbs exponential expression (Hristopulos 2003, 2020a), where  $\mathbf{v}$  is a set of model parameters

$$f_{\mathbf{x}}(\mathbf{x}; \mathbf{v}) = \frac{e^{-H(\mathbf{x}; \mathbf{v})}}{Z(\mathbf{v})} \tag{15}$$

The constant  $Z(\mathbf{v})$  is a normalization factor of the PDF which is obtained by integrating  $\exp[-H(\mathbf{x}; \mathbf{v})]$  over all possible values of the data vector  $\mathbf{x}$ . However, the value of  $Z(\mathbf{v})$  is not needed in spatial prediction.

## Precision matrix formulation

The SLI energy functional  $H(\mathbf{x};\mathbf{v})$ , where  $\mathbf{v}$  is the vector of SLI parameters (to be defined in the following), can be expressed in terms of the precision matrix as follows:

$$H(\mathbf{x};\mathbf{v}) = \frac{1}{2}(\mathbf{x} - \mathbf{m})^\top \mathbf{J}(\mathbf{v}')(\mathbf{x} - \mathbf{m}) \quad (16)$$

In Eq. (16),  $\mathbf{v}' = (c_1, \lambda, \mu, k)^\top$  is a reduced parameter vector and  $\mathbf{J}(\mathbf{v}')$  is a symmetric precision matrix given by the following equation

$$\mathbf{J}(\mathbf{v}') = \frac{1}{\lambda} \left[ \frac{\mathbf{I}}{N} + c_1 \mathbf{J}_1(\mathbf{h}) \right] \quad (17)$$

where  $\lambda, c_1 > 0$  are SLI parameters,  $\mathbf{I}$  is the identity matrix,  $\mathbf{h}$  is the vector of kernel bandwidths, and  $\mathbf{J}_1$  is the interaction network matrix (gradient precision submatrix); the latter is determined by the sampling pattern, the kernel function, and the kernel bandwidth as explained in Hristopulos (2015); Hristopulos et al. (2021).

The SLI vector  $\mathbf{v} = (m_X, c_1, \lambda, \mu, k)^\top$  contains the following parameters:  $m_X$  is the mean value (assumed constant),  $c_1$  controls the amplitude of the term which “mimics” the sum of square gradients, and  $\lambda$  controls the overall amplitude of the fluctuations. Finally,  $\mu > 1$  and  $k \in \{2, 3, 4, \dots\}$  are parameters that control the bandwidth vector  $\mathbf{h} = (h_1, \dots, h_N)^\top$ . The integer parameter  $k$  determines the order of the near-neighbor used to define the local bandwidth, while  $\mu$  is a global multiplication factor. The bandwidth  $h_q$  for a data point  $\mathbf{s}_q$  depends on the distance of  $\mathbf{s}_q$  from its  $k$ -nearest neighbor  $\mathbf{s}_{k(q)}$  as follows

$$h_q = \mu \|\mathbf{s}_q - \mathbf{s}_{k(q)}\|, \quad q = 1, \dots, N \quad (18)$$

In SLI estimation,  $k$  is preselected, while  $m_X, \lambda, c_1, \mu$  are inferred from the data by means of MLE. If a superposition of basis functions is used to model the trend instead of a constant  $m_X$ , the coefficients of the superposition can also be estimated by means of MLE. Such a trend analysis is not used in the current application of SLI.

## SLI prediction

In order to predict the water level at an unmeasured point  $\mathbf{s}_p$  (in this section  $\mathbf{s}_p$  is used instead of  $\mathbf{u}$  for notational convenience), it is inserted in the energy functional  $H$  so that it interacts with the neighboring sampling points via precision matrix elements  $J_{p,n}$ , for  $n = 1, \dots, N$ . The “updated” energy function  $H(\mathbf{x}, x_p; \mathbf{v})$  now defines, in terms of the Boltzmann-Gibbs PDF, the marginal conditional distribution at  $\mathbf{s}_p$ . Due to the local nature of the interactions, most of the  $J_{p,n}$  values are zero. The mode of the conditional

distribution at  $\mathbf{s}_p$  is then given by the following predictive equation (Hristopulos 2015; Hristopulos et al. 2021)

$$\hat{x}_{\text{SLI}}(\mathbf{s}_p) = m_X - \frac{\sum_{q=1}^N J_{q,p}(\mathbf{v})(x_q - m_X)}{J_{p,p}} \quad (19)$$

where  $\hat{x}_{\text{SLI}}(\mathbf{s}_p)$  is the prediction,  $x_q - m_X$  is the fluctuation at location  $\mathbf{s}_q$  and  $J_{q,p}, J_{p,p}$  represent elements of the precision matrix. Note that the SLI prediction of Eq. (19) is independent of the scale coefficient  $\lambda$ .

The precision matrix uses kernel-based weights  $w_{p,q}$  to embody interactions between nearby points. More specifically, there are two weights for each pair of points  $\mathbf{s}_n$  and  $\mathbf{s}_m$  as follows

$$w_{n,m}(\mathbf{h}) = \frac{K(\mathbf{d}_{n,m}/h_n)}{\sum_{l=1}^N \sum_{q=1}^N K(\mathbf{d}_{l,q}/h_l)}, \quad \text{for } n, m = 1 \dots, N \quad (20)$$

$$w_{m,n}(\mathbf{h}) = \frac{K(\mathbf{d}_{n,m}/h_m)}{\sum_{l=1}^N \sum_{q=1}^N K(\mathbf{d}_{l,q}/h_l)}, \quad \text{for } n, m = 1 \dots, N \quad (21)$$

where  $\mathbf{d}_{n,m} = \|\mathbf{s}_n - \mathbf{s}_m\|$  is the Euclidean distance between the points. The bandwidths  $h_n, h_m$  are determined from the sampling density in the neighborhoods of  $\mathbf{s}_n$  and  $\mathbf{s}_m$  respectively. Since the sampling patterns in the neighborhoods of  $\mathbf{s}_n$  and  $\mathbf{s}_m$  can be quite different, in general it holds that  $w_{m,n}(\mathbf{h}) \neq w_{n,m}(\mathbf{h})$ .

The entries of the gradient precision submatrix are determined from the weights as follows (Hristopulos 2015; Hristopulos et al. 2021)

$$[\mathbf{J}_1(\mathbf{h})]_{n,m} = -w_{n,m}(\mathbf{h}) - w_{m,n}(\mathbf{h}) + \delta_{n,m} \sum_{l=1}^N [w_{n,l}(\mathbf{h}) + w_{l,n}(\mathbf{h})] \quad (22)$$

where  $\delta_{n,m} = 1$  if  $n = m$  and  $\delta_{n,m} = 0$  if  $n \neq m$  is the Kronecker delta.

The SLI conditional variance at  $\mathbf{s}_p$  is given approximately by Hristopulos et al. (2021)

$$\sigma_p^2 \approx \frac{\lambda}{c_1} \left[ \sum_{q=1}^N (w_{q,p} + w_{p,q}) \right]^{-1} \quad (23)$$

As evidenced in Eq. (23), higher values of  $c_1$  tend to reduce the variance. This happens due to increasing energy cost for large increments in the energy functional  $H(\mathbf{x};\mathbf{v})$ . In addition, prediction points with more neighbors (points  $\mathbf{s}_q$  such that  $w_{q,p} > 0$  or  $w_{p,q} > 0$ ) have lower conditional variance than prediction points in sparsely sampled regions. Similar to the Kriging variance, the SLI conditional variance does not explicitly depend on the data values. The dependence is

indirect, to the extent that the data influence the estimation of the model parameters.

## Gaussian anamorphosis

In many cases, spatial data follow a skewed PDF that deviates significantly from the normal distribution. In the present study, the water level distribution is clearly asymmetric (see section “[Data description and preprocessing](#)”). Nonlinear, monotonic transformations such as the logarithm, Box-Cox, normal-score, modified Box-Cox versions, and deformed logarithmic transforms are used to restore normality (Box and Cox 1964; Varouchakis et al. 2012; Goovaerts 1997; Hristopulos 2020a; Hristopulos and Baxevani 2022). The application of normalizing transforms is often referred to as “Gaussian anamorphosis”. The spatial analysis is then carried out using the transformed data. Finally, predictions in the original domain are derived by applying the inverse transformation.

In this study, Gaussian anamorphosis is performed using KCDE (see section “[Kernel functions](#)”) coupled with the normal scores transform. KCDE derives  $\hat{F}_K(x)$  from the water level data  $\{x_1, \dots, x_N\}$ , using kernel smoothing to avoid the discontinuities of the staircase CDF function. Normalized estimates,  $\{x_1^*, \dots, x_N^*\}$ , are then calculated based on the normal scores transform. These conform with the standard normal distribution  $\mathcal{N}(0, 1)$ . Spatial interpolation of the normal scores leads to predictions  $\{x_1^*, \dots, x_p^*\}$ . Finally, an inverse transform is used to map the normalized predictions  $\{\hat{x}_1^*, \dots, \hat{x}_p^*\}$  back to water level values  $\{\hat{x}_1, \dots, \hat{x}_p\}$ .

The inverse transform in Gaussian anamorphosis is implemented via a lookup table that maps the predictions of the normalized values back to water level values. The inverse transform is denoted by  $x = \hat{F}_K^{-1}(\Phi(x^*))$ , where  $\Phi(\cdot)$  is the CDF of the standard normal distribution. Hence, the lookup table implements the mapping  $x^* \rightarrow x$ .

To create the lookup table, the expression Eq. (13) for  $\hat{F}_K(x)$  is used to calculate the CDF for a large set of discretization points  $\{z_l\}_{l=1}^L$  (herein,  $L = 10,000$ ). The table thus generated comprises  $L$  rows and three columns that correspond to the discretization points  $\{z_l\}_{l=1}^L$ , the respective normal scores, and the corresponding CDF values  $\{\hat{F}_K(z_l)\}$ . The discretization points are linearly distributed between the minimum and maximum possible values, i.e.,  $z_{\min} = x_{\min} - h$  and  $z_{\max} = x_{\max} + h$  where  $x_{\min} = \min\{x_1, \dots, x_N\}$  and  $x_{\max} = \max\{x_1, \dots, x_N\}$  as shown in Pavlides et al. (2022). The inverse transform of any prediction  $\hat{x}^*$  is calculated as follows: first, the respective  $\Phi(\hat{x}^*)$  is determined; next, the closest probability level to  $\Phi(\hat{x}^*)$  in the lookup table is determined; if this corresponds to  $\hat{F}_K(z_o)$ , where  $z_o \in \{z_1, \dots, z_L\}$ , the inverse transform of  $\hat{x}^*$  is given by  $\hat{x} = z_o$ .

## Assessment of predictions

Validation measures are used to assess the quality of the spatial prediction achieved by the models. Furthermore, stochastic methods provide estimates of the prediction uncertainty for each location.

The predictive accuracy of both UK and SLI was assessed using LOOCV (Chilès and Delfiner 2012). In LOOCV the sampling sites  $\mathbf{s}_n$ ,  $n = 1, \dots, N = 72$  and the corresponding measurements are cyclically removed from the data one at a time, and the water level,  $\hat{x}_n$  at the point removed is estimated from the remaining 71 drillholes. The estimation error  $\epsilon_n = x_n - \hat{x}(\mathbf{s}_n)$ , for  $n = 1, \dots, 72$  is then calculated.

Ninety percent prediction intervals (i.e., based on a 90% confidence level) are used to characterize the uncertainty of the predictions. These intervals are calculated in two steps. First, prediction intervals for each point are generated for the transformed (normalized) data based on the prediction variance; the latter is given by Eq. (12) for UK and Eq. (23) for SLI. The 90% prediction interval at location  $\mathbf{u}$  for either method is given by

$$[\hat{x}^*(\mathbf{u}) - 1.645\sigma(\mathbf{u}), \hat{x}^*(\mathbf{u}) + 1.645\sigma(\mathbf{u})] \quad (24)$$

where  $\hat{x}_{0.05}^*(\mathbf{u}) = \hat{x}^*(\mathbf{u}) - 1.645\sigma(\mathbf{u})$  is the 5% quantile, and  $\hat{x}_{0.95}^*(\mathbf{u}) = \hat{x}^*(\mathbf{u}) + 1.645\sigma(\mathbf{u})$  is the 95% quantile of the Gaussian predictive distribution at  $\mathbf{u}$ . The water level prediction intervals at  $\mathbf{u}$  are obtained by applying the inverse of the normalizing transformation to the 5% and 95% quantiles,  $\hat{x}_{0.05}^*(\mathbf{u})$  and  $\hat{x}_{0.95}^*(\mathbf{u})$  respectively. The inversion employs the principle of quantile invariance under monotonic transformations (Hristopulos 2020a) and is implemented by means of the lookup table defined in section “[Gaussian anamorphosis](#)”. Thus, the respective water level quantiles  $\hat{x}_{0.05}(\mathbf{u})$  and  $\hat{x}_{0.95}(\mathbf{u})$  are obtained. To better visualize the uncertainty of the prediction, the 90% prediction interval range  $R = \hat{x}_{0.95}(\mathbf{u}) - \hat{x}_{0.05}(\mathbf{u})$  is displayed.

## Data description and preprocessing

The study area comprises three mines operating in the same region in northern Greece. The area is geologically composed of various metamorphic and igneous rocks; biotite gneisses with intercalations of marble horizons are located in the area; granodiorites and acid to intermediate subvolcanic rocks intrusions into the gneisses cover small areas. The study area is characterized hydrogeologically as semipermeable, and its vertical profile comprises three hydrostratigraphic units. Secondary porosity (fractures) is responsible for hydrological connectivity. In certain areas, the rocks exhibit strong fragmentation which increases with

depth thus creating pathways for slow groundwater flow (ENVECO 2010). The data used in this study comprise the 10-year average (2011–2020) of biannual water level measurements (below surface, BSL), from 72 drillholes. All the values are negative since water level is measured in terms of meters below surface. The measurement boreholes are located around the three mines operating in the area of interest (see Fig. 1b):

- Mine A (28 drillholes)
- Mine B (28 drillholes)
- Mine C (16 drillholes)

### Preliminary data analysis

This section presents a statistical summary of the *temporally averaged* water levels from the entire area. The frequency histogram and the data locations are shown in Fig. 1. The median distance of a borehole to its second-nearest neighbor ( $k = 2$ ) is 488.15 m. Table 2 shows the summary statistics of the data.

Based on the skewness value in Table 2 and the shape of the histogram, the data deviate strongly from the normal (Gaussian) distribution. As discussed in section “Kriging”, the Kriging variance does not adequately represent the prediction uncertainty for non-Gaussian data. To address this issue, Gaussian anamorphosis is employed in order to normalize the data (see section “Gaussian anamorphosis”).

### Gaussian anamorphosis of water level data

In order to address the non-Gaussian distribution of the data, the data-driven Gaussian anamorphosis method is

**Table 2** Summary water level data statistics. *SD* standard deviation

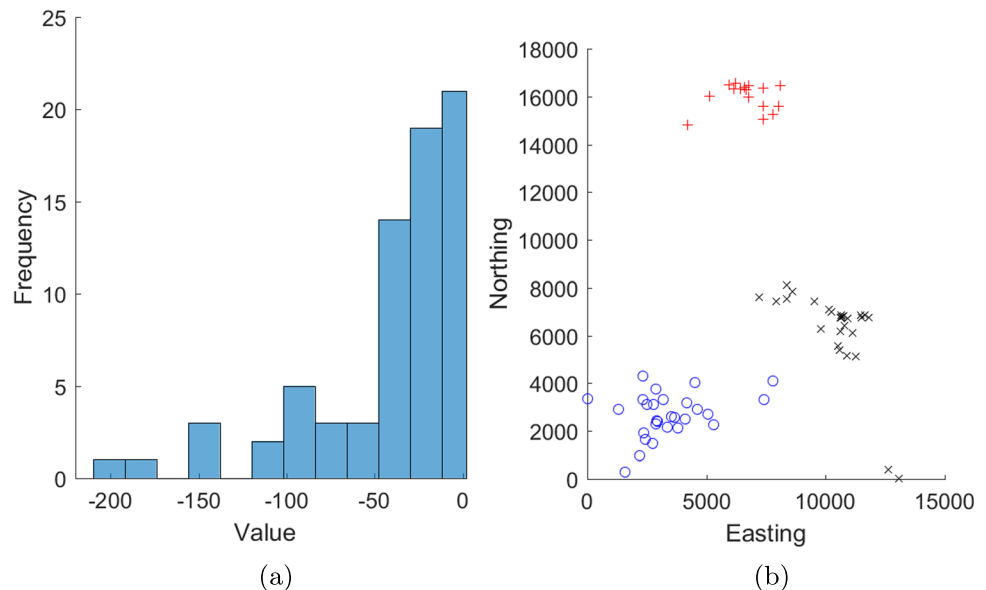
| Minimum | Maximum | Mean   | Median | SD    | Skewness |
|---------|---------|--------|--------|-------|----------|
| −208    | −1      | −41.82 | −28    | 45.57 | −1.75    |

The sample contains 72 data points that represent the average of biannual measurements over the 10-year period at each sampling site. All statistics are measured in meters below sea level, except for skewness (dimensionless measure)

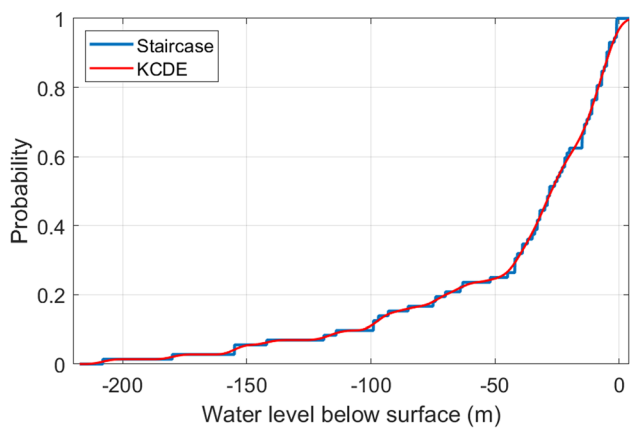
employed. The method is described in section “Gaussian anamorphosis”. First, KCDE is applied with the triweight kernel (see Table 1) to provide a nonparametric CDF,  $\hat{F}_K(x)$ , as discussed in Pavlides et al. (2022). The plug-in kernel bandwidth was calculated at  $h = 9.33\text{m}$  ( $h$  in KCDE refers to water level, not geographical distance).

Figure 2 shows the KCDE-derived  $\hat{F}_K(x)$  in comparison to the staircase  $\hat{F}_s(x)$ . In contrast with the latter,  $\hat{F}_K(x)$  is a continuous curve which was constructed using the lookup table. Figure 3 shows the histogram and the normal probability plot for the transformed data  $\{x_1^*, \dots, x_N^*\}$  which are obtained after normalization. As is evident in these plots, the transformed data  $\{x_1^*, \dots, x_N^*\}$  conform to the standard normal distribution  $\mathcal{N}(0, 1)$ . Spatial prediction is carried out using the normalized data. The predictions  $\{\hat{x}_1^*, \dots, \hat{x}_p^*\}$  obtained at the nodes of the map grid  $\mathbb{G}$  are transformed back to water level values  $\{\hat{x}_1, \dots, \hat{x}_p\}$  with the help of the lookup table as explained in section “Gaussian anamorphosis”.

**Fig. 1** **a** Frequency histogram of the 72 water level measurements (m below surface). **b** drillhole locations for the three mines: mine A (blue o), mine B (black x), mine C (red +). Coordinates are in meters







**Fig. 2** Empirical staircase CDF,  $\hat{F}_s(x)$  (blue online) and KCDE-derived continuous CDF,  $\hat{F}_K(x)$  based on Eq. (13) (red online). The triweight kernel with bandwidth  $h = 9.33$  is used in KCDE

### Results: universal Kriging

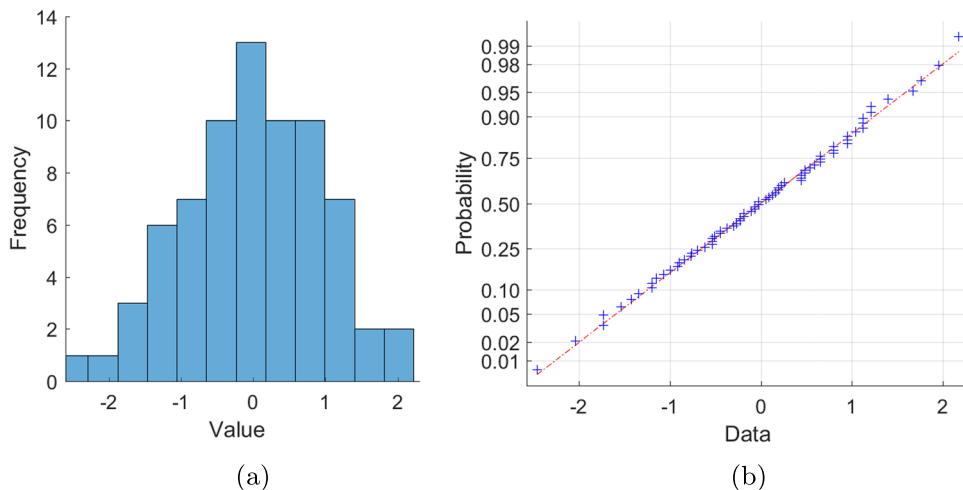
The parameters of the spherical covariance model for the normalized data  $\{x_1^*, \dots, x_N^*\}$  are obtained by maximizing the likelihood Eq. (11). The spherical covariance is obtained from the respective variogram model of Eq. (5) by means of Eq. (3). The MLE initialization values for the variogram parameters are  $\sigma_0^2 = 1.15\text{m}$ ,  $\xi_0 = 2800\text{m}$ ,  $c_{o;0} = 0.11$ , where  $c_{o;0}$  is the nugget effect. The initialization parameters for the variogram have been estimated by fitting the spherical model to the experimental variogram. The initial values for the trend parameters (obtained by means of linear regression) are  $a_{0;0} = -0.93$ ,  $a_{1;0} = 7.3 \cdot 10^{-5}$ , and  $a_{2;0} = 5.9 \cdot 10^{-5}$ . The optimal MLE parameters for the variogram model (spherical and nugget) are  $\hat{\sigma}^2 = 0.89$ ,  $\hat{\xi} = 1,653 \text{ m}$ , and  $c_o = 0.106$ . The optimal variogram model (based on the MLE parameters) and

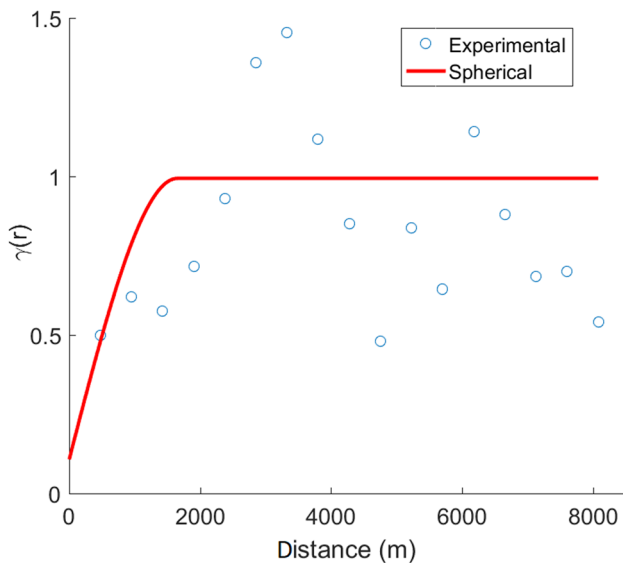
the experimental variogram are compared in Fig. 4. The experimental variogram is calculated for the residuals which are obtained by subtracting from the data the MLE linear trend with coefficients  $a_{0;0} = -0.63$ ,  $a_{1;0} = 3.98 \cdot 10^{-5}$ , and  $a_{2;0} = 5.00 \cdot 10^{-5}$ .

UK is then applied (using the estimated parameters) to interpolate the transformed data values  $\{x_1^*, \dots, x_N^*\}$  using Eq. (8). Estimates  $\{\hat{x}_1^*, \dots, \hat{x}_p^*\}$  are generated on a grid with  $112 \times 89$  cells (corresponding to a cell size of  $150 \times 150 \text{ m}$ ). The range of the Kriging search neighborhood was selected at 2,400 m in order to provide adequate coverage of the mining areas and the minimum number of neighbors was set at  $K = 2$ . Estimates were generated only for points  $\mathbf{u}$  with search neighborhoods containing at least three points in the sampling set.

The optimal prediction  $\hat{x}_{\text{SLI}}(\mathbf{u})$ , as well as the estimates of the quantiles  $\hat{x}_{0.05}(\mathbf{u})$  and  $\hat{x}_{0.95}(\mathbf{u})$  are presented in Figs 5 and 6. The uncertainty of the predictions is visualized by means of the 90% prediction interval range  $R_{\text{UK};90} = \hat{x}_{0.95}(\mathbf{u}) - \hat{x}_{0.05}(\mathbf{u})$  shown in Fig 5b. High values of  $R_{\text{UK};90}(\mathbf{u})$  at location  $\mathbf{u}$  denote higher uncertainty of the prediction. Note that the predictive distribution of  $X(\mathbf{u})$  is not necessarily normal due to the application of the inverse transform. Hence,  $\hat{x}_{0.05}(\mathbf{u})$  and  $\hat{x}_{0.95}(\mathbf{u})$  are not symmetrically placed around  $\hat{x}_{\text{UK}}(\mathbf{u})$  as is evident in Fig. 6. For example, the prediction at  $\mathbf{u} = (10200, 5400)$  is  $\hat{x}(\mathbf{u}) = -18 \text{ m}$  and  $R_{\text{UK};90}(\mathbf{u}) = 76 \text{ m}$  while  $\hat{x}_{0.05}(\mathbf{u}) = -78 \text{ m}$  and  $\hat{x}_{0.95}(\mathbf{u}) = -2 \text{ m}$ . Thus, there is 90% probability that  $X(\mathbf{u}) \in [-78 \text{ m}, -2 \text{ m}]$ , but the UK prediction is not given by the interval's mean (i.e.,  $-36 \text{ m}$ ) but rather by  $\hat{x}(\mathbf{u}) = -18\text{m}$  which corresponds to the median of the predictive distribution.

**Fig. 3 a** Histogram of the normalized average (over 10 years of biannual measurements) water level at 72 sites. **b** Normal probability plot for the normalized data





**Fig. 4** Variogram of the residuals (determined by removing the MLE-based linear trend). The spherical model is shown with a continuous curve (red line), and the experimental variogram is shown by circle markers (blue circle)

**UK cross-validation**

The UK predictive accuracy was assessed using LOOCV (section “Assessment of predictions”). In LOOCV sampling sites  $s_n, n = 1, \dots, N = 72$  and the corresponding measurements are cyclically removed from the data, and the value of the water level,  $\hat{x}_n$  at the point removed is estimated from the remaining 71 drillholes. The estimation error  $\epsilon_n = x_n - \hat{x}(s_n)$ , for  $n = 1, \dots, 72$  is calculated.

The histogram of UK estimation errors for all mines in the study area is shown in Fig. 7. Standard validation metrics are shown in Table 3. The values of the data range from  $x_{\max} = -1\text{m}$  to  $x_{\min} = -208\text{m}$ . Most of the validation errors

are in  $[-50, 50]\text{m}$ , with the exception of a few locations that return errors greater than 50 m (in magnitude). The presence of few high validation errors is not surprising considering the small sample size and the significant spatial dispersion of the sampling points. These factors lead to large prediction intervals (as shown in the uncertainty map of Fig 5b). Taking into account the preceding factors, the UK interpolation performance is considered adequate.

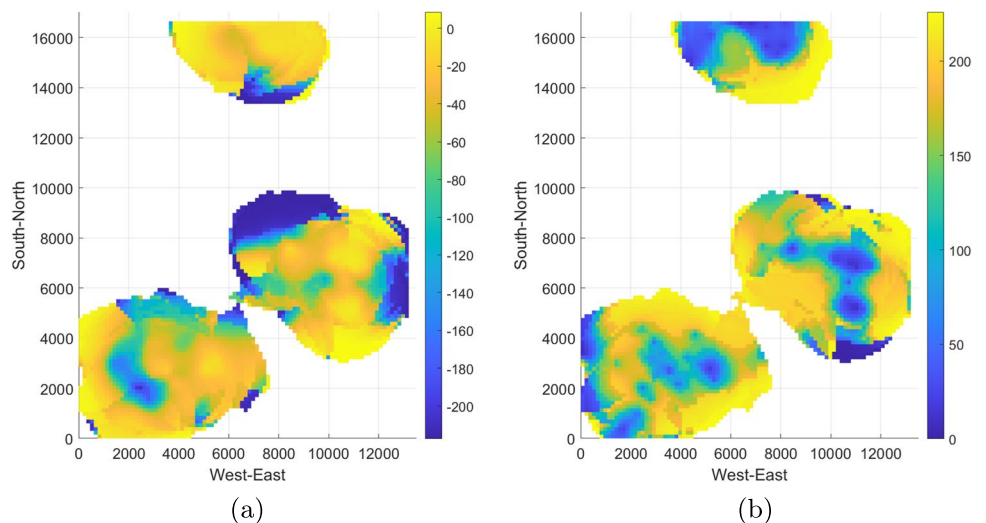
**Results: SLI**

The water level in the study area is estimated using SLI (see section “Stochastic local interaction model”) with the quadratic kernel (see Table 1). The transformed data values  $\{x_1^*, \dots, x_N^*\}$  are used for SLI interpolation and the predictions are subsequently back-transformed back to meters BSL. The SLI model parameters  $c_1, \lambda, \mu$  and  $m_X$  are estimated by MLE. The initialization values for the MLE are:  $\mathbf{v}'_0 = (c_1 = 2.5, \lambda = 0.1, \mu = 1.25, m_X = -0.015)^\top$ . The SLI model parameters estimated by MLE are  $\mathbf{v}' = (c_1 = 0.926, \lambda = 0.032, \mu = 4.18, m_X = -0.015)^\top$ .

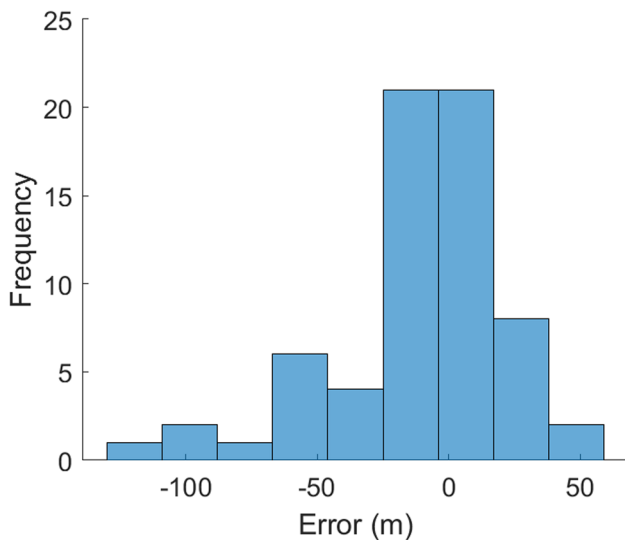
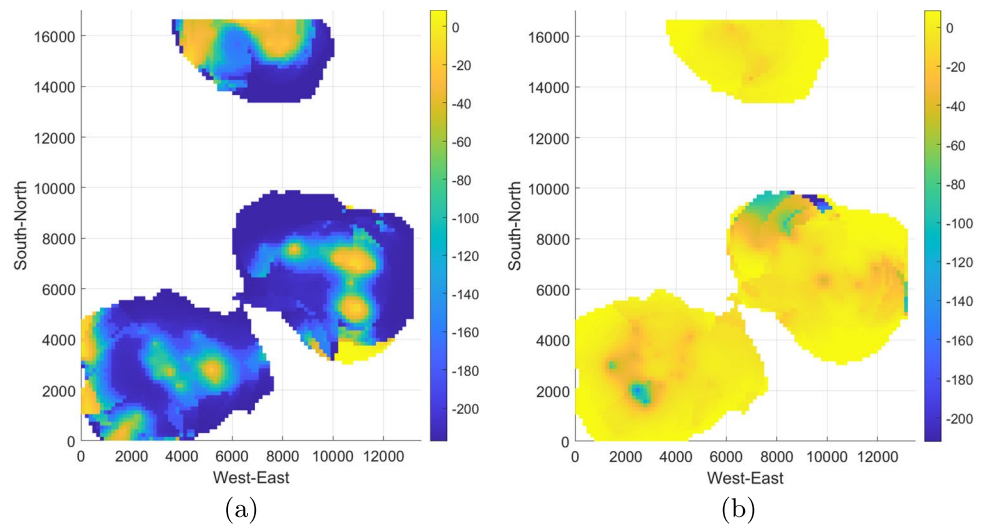
Using the Eqs. (19) and (23), the optimal SLI predictions  $\hat{x}_{\text{SLI}}^*(\mathbf{u})$  and respective variances  $\sigma_{\text{SLI}}^2(\mathbf{u})$  are estimated. The same map grid  $\mathbb{G}$  is used for UK (see section “Results: universal Kriging”). In contrast with Kriging, the search neighborhood in SLI is automatically adjusted by the algorithm through the parameter  $\mu$  and the bandwidth vector  $\mathbf{h}$  instead of being set by the user. For comparison purposes, the SLI predictions are used only at the map-grid nodes where UK predictions are available. The normality of the transformed data  $\{x_1^*, \dots, x_N^*\}$  allows using the standard deviation of the SLI predictions to determine prediction intervals using the method presented in section “Assessment of predictions”.

The results are presented in Figs. 8 and 9.

**Fig. 5 a** Map of estimated water level (m below surface) for the three mines in the study area. **b** Range of the 90% water level prediction intervals; high values denote higher uncertainty. Coordinates are in meters



**Fig. 6** UK-based maps of the lower and upper limits of 90% prediction intervals: **a** lower quantile  $\hat{x}_{0.05}(\mathbf{u})$ ; **b** upper quantile  $\hat{x}_{0.95}(\mathbf{u})$ . Coordinates are in meters



**Fig. 7** Histogram of the UK validation errors for the three mines in the study area

**Table 3** Validation measures (LOOCV) for assessment of UK interpolation performance

| $\rho$ | ME (m) | MAE (m) | MaxAE (m) | RMSE (m) |
|--------|--------|---------|-----------|----------|
| 0.697  | -10.27 | 22.82   | 125.64    | 34.29    |

$\rho$  Pearson’s correlation coefficient, *ME* mean error, *MAE* mean absolute error, *MaxAE* maximum absolute error, *RMSE* root mean square error

The SLI algorithm returns large search neighborhoods as implied by  $\mu = 4.18$  (see section “Stochastic local interaction model”): the bandwidth Eq. (18) dictates that the local bandwidth  $h_u$  at the point  $\mathbf{u}$  is the distance to the  $k$ -nearest neighbor ( $k = 2$ ) multiplied by  $\mu$ . As shown in Fig. 1,

$k$ -nearest neighbor distances vary from small (43 m) to large (2,977 m).

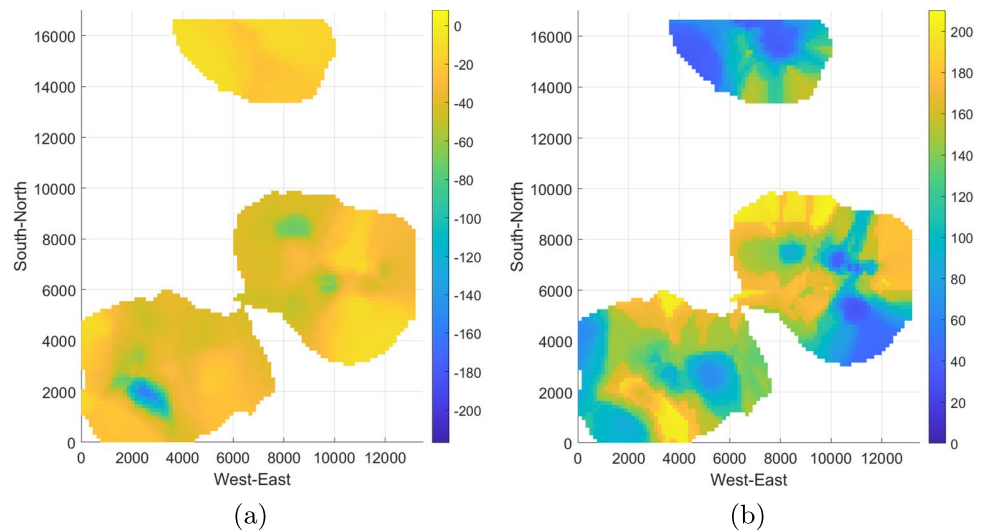
As is evident in Figs. 5b and 8b, the SLI prediction interval range has a minimum value,  $\min_{\mathbf{u} \in G} \{R_{SLI;90}(\mathbf{u})\}$ , of 31.9 m. For UK, the prediction interval has a lower minimum, i.e.,  $\min_{\mathbf{u} \in G} \{R_{UK;90}(\mathbf{u})\} = 0.7$  m. The reason is that UK explicitly minimizes the prediction variance, and thus it generates a smaller range for prediction intervals.

### SLI application to mine B

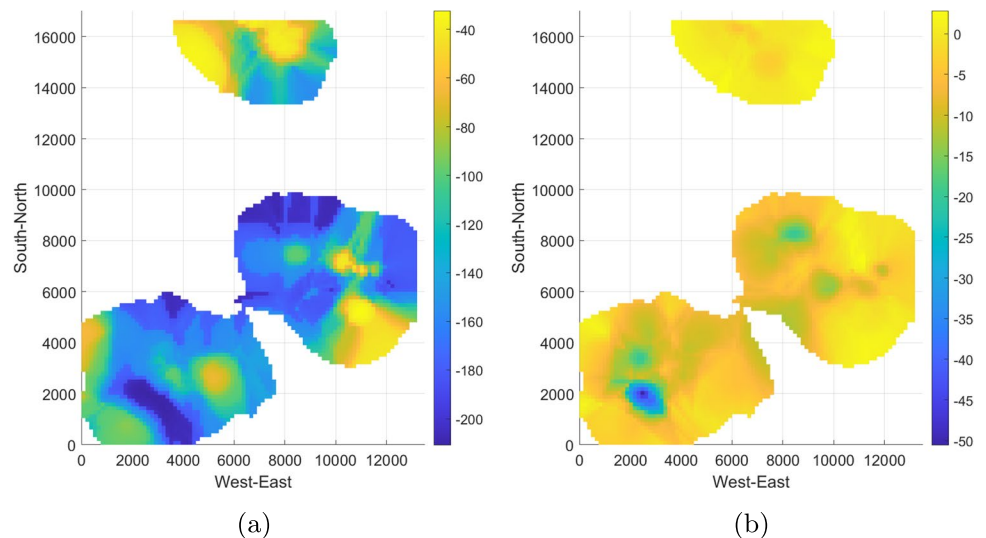
Stochastic local interaction can handle large datasets due to the sparse structure of the model (Hristopulos 2015; Hristopulos et al. 2021). However, the locally adaptive nature of the SLI predictor also allows model estimation and prediction even for small samples. In this section SLI is applied to the dataset for mine B. This mine involves 28 sampling points, a sample size which is considered small for most stochastic methods. The MLE initialization values for mine B are the same as for the entire area:  $\mathbf{v}_0 = (c_1 = 2.5, \lambda = 0.1, \mu = 1.25, m_X = -0.56)^T$ . The optimal SLI parameters obtained by MLE are:  $\mathbf{v} = (c_1 = 2.881, \lambda = 0.046, \mu = 0.938, m_X = 0.163)^T$ .

The maps for  $\hat{x}_{SLI}(\mathbf{u})$ ,  $\hat{x}_{0.05}(\mathbf{u})$ ,  $\hat{x}_{0.95}(\mathbf{u})$  and the range  $R_{SLI;90}$  are shown in Fig. 10. The prediction of the water level values using only data from mine B is similar to the SLI implementation using the data from the entire area. The water level in the northwestern area of the mine is lower than in the southeast, although SLI for mine B gives slightly higher predictions for the low water level area in the northeast than SLI for the entire area. The prediction uncertainty using only data from mine B follows a similar spatial distribution as the interval shown in Fig. 8b. However, the uncertainty is lower in the northwestern area and the eastern part of the mine than the SLI implementation using all the data, albeit it is still relatively high.

**Fig. 8** **a** SLI prediction map of water level (m below surface level) for the three mines. **b** Range of water level 90% prediction intervals. Coordinates are in meters



**Fig. 9** SLI-based maps of the lower and upper limits of 90% prediction intervals: **a** lower quantile  $\hat{x}_{0.05}(\mathbf{u})$ ; **b** upper quantile  $\hat{x}_{0.95}(\mathbf{u})$ . Coordinates are in meters



The differences for the mine B results between Figs. 8–9 and Fig. 10 are expected since the maps of Fig. 10 are produced with fewer data that determine a different SLI model.

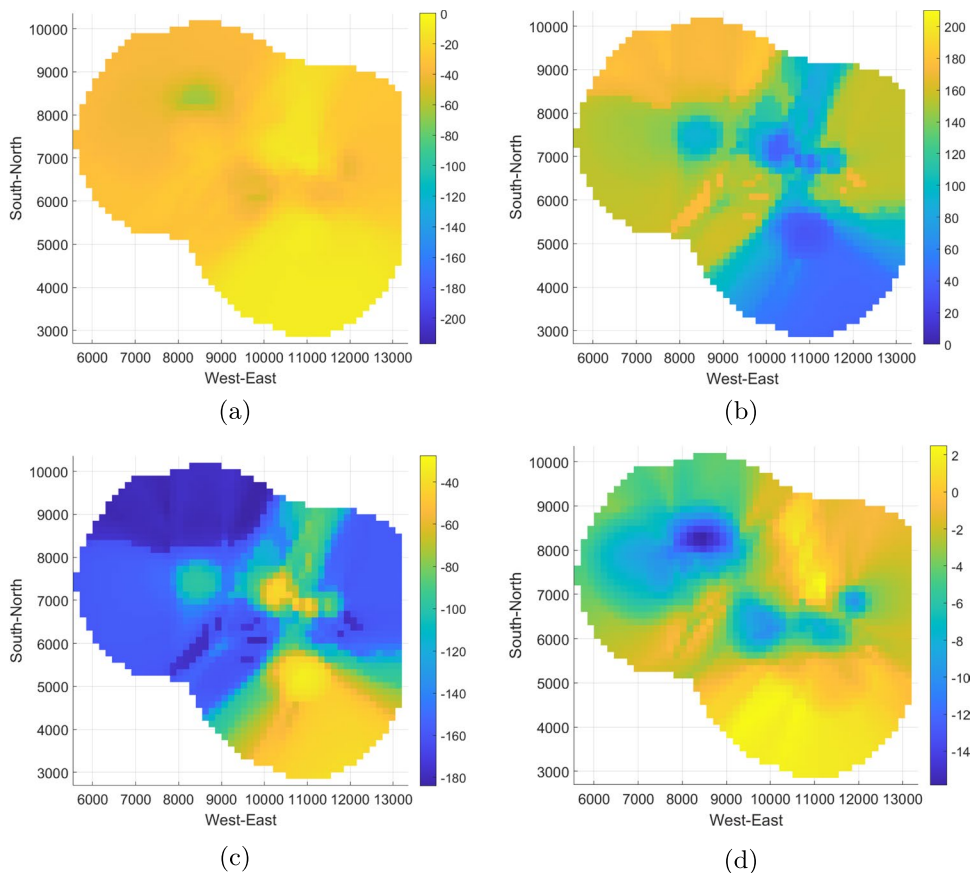
### SLI cross-validation

Leave-one-out cross-validation is used to assess the SLI predictive performance, first for the SLI model based on the entire area and then for the SLI model obtained for mine B. The validation metrics for the SLI estimation are shown in Table 4. The histograms of the errors are shown in Fig. 11. The data values for the entire area range between  $x_{\max} = -1$  m to  $x_{\min} = -208$  m. However, the lowest LOOCV prediction is  $\min_{i=\{1,2,\dots,N\}} \hat{x}(s_i) = -121.9$  m, i.e., the minimum is significantly overestimated. One reason for this behavior is the difficulty of estimating the minimum value, once it is removed, in a sparsely sampled

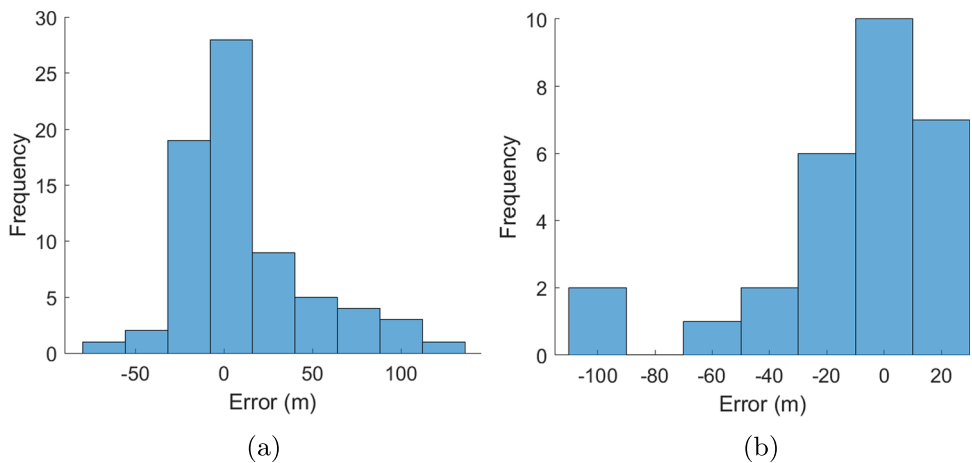
area; this effect is similar to the Kriging smoothing effect (Yamamoto 2005; Pavlides et al. 2015). Another reason is that a trend function in the SLI model was not used. For mine B, the values of the data range from  $x_{\max} = -1$  m to  $x_{\min} = -142$  m. SLI again overestimated the lowest value yielding a minimum of  $-39.2$  m.

For the SLI using data from the entire area, there are a few large errors ( $\epsilon > 50$  m). For SLI predictions using only mine B data, there are three locations that return high validation errors. In both cases, the SLI overestimates the water level in both mine B and mine C. However, considering the range of data values, the clustering of the data locations and the small number of data points (especially for mine B), and the LOOCV measures over the study areas (Table 4), the SLI performance is considered adequate overall.

**Fig. 10** SLI maps for mine B: **a** predicted water levels (m below surface); **b** range of the 90% water level prediction intervals; **c** lower quantile  $\hat{x}_{0.05}(\mathbf{u})$ ; **d** higher quantile  $\hat{x}_{0.95}(\mathbf{u})$ . Coordinates are in meters



**Fig. 11** Histograms of SLI leave-one-out cross-validation errors: **a** entire area, **b** mine B



**Table 4** Validation measures for the SLI estimations

| Area      | $\rho$ | ME (m) | MAE (m) | MaxAE (m) | RMSE (m) |
|-----------|--------|--------|---------|-----------|----------|
| All mines | 0.620  | -12.06 | 25.47   | 136.28    | 38.17    |
| Mine B    | 0.499  | -12.01 | 22.70   | 105.44    | 34.57    |

$\rho$  Pearson’s correlation coefficient, ME is the mean error, MAE is the mean absolute error, MaxAE is the maximum absolute error and RMSE is the Root Mean Square Error

### Discussion

This section provides a comparative discussion of the UK and SLI cross validation results. For ease of reference, the validation measures are gathered in Table 5. In the case of UK, the validation measures for Mine B are derived using the variogram model estimated from the entire area, as 28

**Table 5** Leave-one-out cross validation measures for SLI and UK

| Area            | $\rho$ | ME (m) | MAE (m) | MaxAE (m) | RMSE (m) |
|-----------------|--------|--------|---------|-----------|----------|
| SLI entire area | 0.620  | -12.06 | 25.47   | 136.28    | 38.17    |
| UK entire area  | 0.697  | -10.27 | 22.82   | 125.64    | 34.29    |
| SLI mine B      | 0.499  | -12.01 | 22.70   | 105.44    | 34.57    |
| UK mine B       | 0.396  | -10.24 | 22.87   | 106.39    | 35.28    |

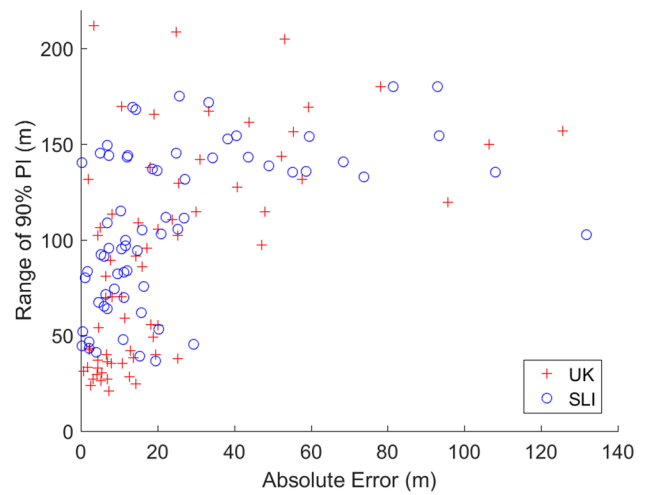
$\rho$  Pearson’s correlation coefficient, *ME* mean error, *MAE* mean absolute error, *MaxAE* maximum absolute error and *RMSE* root mean square error. UK validation measures for mine B use the variogram model inferred from the data over the entire area

data locations are considered too few to reliably estimate the variogram model. The SLI parameters were inferred separately for mine B, and thus a different set of SLI parameters was used (see section “[SLI application to mine B](#)”) than for the entire area.

UK validation measures are slightly better for the entire area. This result is attributed to the fact that UK employs a linear trend Eq. (2) to formulate predictions, while SLI is employed with a constant mean (see Eq. 19). This choice creates a disadvantage for SLI, since the spatial distribution of the data over the entire area does not seem to comply with the stationarity assumption. This shortcoming will be fixed in future implementations of SLI.

Both methods perform adequately given the small sample size, although they significantly overestimate the (few) low values as evidenced in the MaxAE values in Table 5 and the error histograms of Figs. 7 and 11. In addition, both methods give high uncertainties in areas that are distant from the sampling sites. However, these shortcomings point primarily to sampling, not methodological deficiencies.

The differences between SLI and UK are illustrated in Fig. 12. As expected from the results in the previous

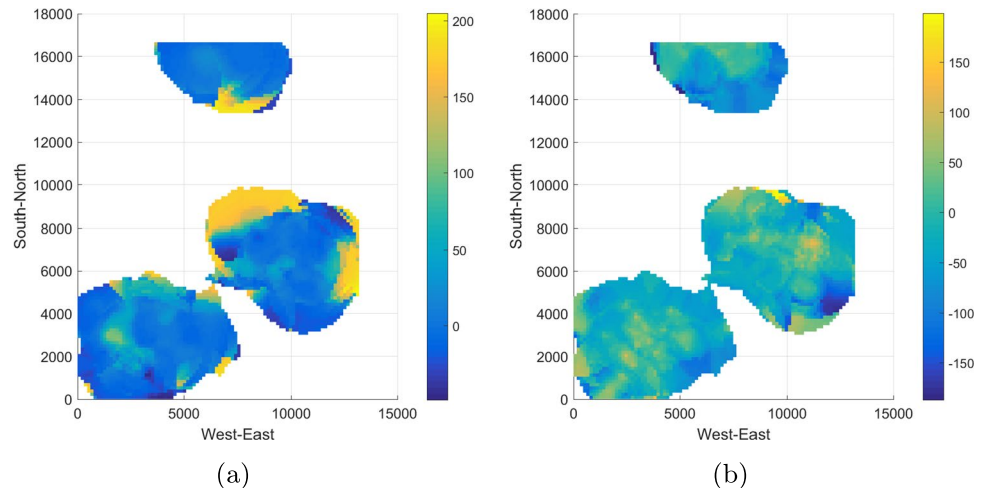


**Fig. 13** Validation absolute errors compared to the range of the 90% water level prediction interval (PI) for SLI (blue circle) and UK (red cross)

sections, the prediction interval (PI) range of SLI,  $R_{SLI}$ , is generally higher than the PI range for UK,  $R_{UK}$ . Kriging methods minimize the prediction variance  $\sigma_{UK}^2$ , which affects the  $R_{UK}$ . SLI prediction is instead based on maximizing the conditional PDF by means of the energy functional  $H(\mathbf{x};\mathbf{v})$  which can lead to higher prediction variances at many prediction points.

SLI gives significantly higher estimates than UK at certain points in mine B. The UK predictions drop significantly in the north of the study domain, i.e., for  $s_2 > 8000$  m (see Fig. 5), while SLI predictions remain high (see Fig. 10). The linear trend in the UK model improves the validation measures compared to a preliminary analysis based on ordinary Kriging. On the other hand, the sudden drop of the water level in the northern section of mine B, forces UK to model a significant linear slope in the trend function, leading to

**Fig. 12** Differences between SLI and UK. **a** Prediction:  $\hat{x}_{SLI}(\mathbf{u}) - \hat{x}_{UK}(\mathbf{u})$  **b** range:  $R_{SLI}(\mathbf{u}) - R_{UK}(\mathbf{u})$ . Coordinates are in meters



decreasing water levels further North. Such predicted drops cannot be verified due to the lack of data. Similar issues are present in a small part of the southern segment of mine C and the western part of mine B. It is impossible to confirm whether SLI or UK predictions are better in these areas without additional data. Figures 6 and 9 show that both UK and SLI predictions are within each other method's prediction intervals.

Overall, in most cases both SLI and UK prediction intervals (based on LOOCV) include the data values at the respective point (six values outside of PI for SLI and four values for UK). For SLI, 33 of the predicted values  $\hat{x}_{SLI}$  are within the interquartile (IQ) range of  $\hat{x}_{0.25}$  to  $\hat{x}_{0.75}$ . UK performs slightly worse with 30 of the predicted values  $\hat{x}_{UK}$  inside the IQ range.

In Fig. 13, the LOOCV absolute error for each of the data sites is plotted against the range of the 90% prediction interval  $R_{SLI}(s)$  for SLI or  $R_{UK}(s)$  for UK. As evidenced in the figure, UK gives both the highest and lowest error values. Furthermore, locations with high absolute error values ( $|x(s) - \hat{x}(s)| > 50$  m) have high PI range ( $R(s) > 100$  m). Thus, for both methods high absolute error values are concentrated in areas with high uncertainty.

In addition, both methods estimate prediction intervals that are wide enough to accommodate the prediction of the other method at most locations. In both cases, the LOOCV results are reasonable given the small size of the dataset. The SLI method allows for constructing a spatial model that can give reasonable predictions even with a dataset as small as that of mine B (28 points).

## Conclusions

In this research, the estimation of groundwater levels in sparsely sampled areas with mining activity is investigated. This case study involves the aggregate water level below surface for three mines in northern Greece. The recently proposed kernel-based KCDE method (Pavlidis et al. 2022) is used to estimate the non-Gaussian distribution of water level and to conduct Gaussian anamorphosis. The predictive performance of two stochastic interpolation approaches is compared, UK and the more recently developed SLI method. SLI is a spatial interpolation method designed to provide computational efficiency for large spatial and spatiotemporal datasets (Hristopulos 2015; Hristopulos and Agou 2020; Hristopulos et al. 2021). While the primary objective of SLI is to efficiently extract correlations from large data, as shown herein, it can also be applied to small datasets that do not allow reliable estimation of the variogram function which is needed for the application of Kriging methods.

The cross-validation measures show evidence of reasonable interpolation performance. However, in certain areas the 90% prediction intervals are rather wide, indicating high prediction uncertainty. UK takes into account a linear trend function and minimizes the prediction variance, thus allowing for tighter prediction ranges than SLI. On the other hand, SLI provides an automatically adaptable search radius, computational efficiency, and the ability to estimate the spatial model even with the very sparse dataset of mine B. As discussed in Section “Discussion”, the introduction of a linear trend model based on a small dataset can be misleading. While UK cross-validation measures are slightly better for the entire area, application of SLI mine B leads to slightly better performance than UK. Furthermore, the trend in areas outside the data clusters leads to rapid decrease of the UK predicted water levels (see Fig. 5) that can not be confirmed based on the available data.

The SLI predictor can be modified to include trend functions as in UK—for example, promising results were obtained by incorporating a polynomial trend in the spatiotemporal SLI predictor (Hristopulos and Agou 2020). Management of water resources is important in areas of mining activity. Future studies could focus on spatiotemporal prediction methodologies in order to allow monitoring seasonal changes of the water level due to mining activities. Furthermore, the KCDE-based estimate of the CDF can be used in conjunction with conditional simulation methods, such as Kriging polarization (Chilès and Delfiner 2012; Hristopulos 2020a; Olea 2012), in order to better capture the spatial variability of water level.

**Funding** Open access funding provided by HEAL-Link Greece.

**Code availability** The MATLAB code used for SLI analysis can be downloaded from the [website](#) of the Geostatistics Laboratory, Technical University of Crete.

## Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abiye T, Masindi K, Mengistu H, Demlie M (2018) Understanding the groundwater-level fluctuations for better management of groundwater resource: a case in the Johannesburg region. *Groundw Sustain Dev* 7:1–7
- Adler RJ (1981) *The geometry of random fields*. Wiley, New York
- Agou VD (2016) Geostatistical analysis of precipitation on the island of Crete. MSc Thesis, Technical University of Crete, Chania, Crete, Greece
- Agou VD, Pavlides A, Hristopulos DT (2022) Spatial modeling of precipitation based on data-driven warping of Gaussian processes. *Entropy* 24(3):321
- An Y, Lu W, Cheng W (2015) Surrogate model application to the identification of optimal groundwater exploitation scheme based on regression kriging method: a case study of Western Jilin province. *Int J Environ Res Public Health* 12(8):8897–8918
- Botev ZI, Grotowski JF, Kroese D (2010) Kernel density estimation via diffusion. *Ann Stat* 38(5):2916–2957
- Box GEP, Cox DR (1964) An analysis of transformations. *J R Stat Soc Ser B Methodol* 26(2):211–252
- Budiman JS, Al-Amri NS, Chaabani A, Elfeki AM (2022) Geostatistical based framework for spatial modeling of groundwater level during dry and wet seasons in an arid region: a case study at Hadat Ash-Sham experimental station, Saudi Arabia. *Stoch Env Res Risk A* 36(8):2085–2099
- Calzolari C, Ungaro F (2012) Predicting shallow water table depth at regional scale from rainfall and soil data. *J Hydrol* 414–415:374–387
- Chilès JP, Delfiner P (2012) *Geostatistics: modeling spatial uncertainty*. Wiley, Hoboken, NJ
- Christakos G (1992) *Random field models in earth sciences*. Academic, San Diego
- Cressie N (1990) The origins of kriging. *Math Geol* 22(3):239–252
- Cressie N (1993) *Spatial statistics*. Wiley, New York
- Dai H, Ren L, Wang M, Xue H (2011) Water distribution extracted from mining subsidence area using kriging interpolation algorithm. *Trans Nonferrous Metals Soc China* 21:s723–s726
- Davies P, Lawrence S, Turnbull J, Rutherford I, Silvester E, Grove J, Macklin MG (2020) Groundwater extraction on the goldfields of Victoria, Australia. *Hydrogeol J* 28(7):2587–2600
- ENVECO (2010) Environmental impact assessment report for mining and metallurgy operations, Appendix 2, map 9-2 (in Greek). ENVECO, Athens
- Evans SW, Jones NL, Williams GP, Ames DP, Nelson EJ (2020) Groundwater level mapping tool: an open source web application for assessing groundwater sustainability. *Environ Model Softw* 131:104782
- Fletcher R (2000) *Practical methods of optimization*, 2nd edn. Wiley, Chichester, England
- Ghosh S (2018) *Kernel smoothing: principles, methods and applications*. Wiley, Hoboken, NJ
- 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
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, New York
- Hristopulos DT (2015) Stochastic local interaction (SLI) model: bridging machine learning and geostatistics. *Comput Geosci* 85:26–37
- Hristopulos DT (2020a) *Random fields for spatial data modeling: a primer for scientists and engineers*. Springer, Dordrecht, The Netherlands
- Hristopulos DT (2020b) Suite of MATLAB programs for the implementation of gradient-based SLI estimation and interpolation.
- Hristopulos DT (2003) Spartan Gibbs random field models for geostatistical applications. *SIAM J Sci Comput* 24(6):2125–2162
- Hristopulos DT, Agou VD (2020) Stochastic local interaction model with sparse precision matrix for space-time interpolation. *Spat Stat* 40:100403
- Hristopulos DT, Baxevani A (2022) Kaniadakis functions beyond statistical mechanics: weakest-link scaling, power-law tails, and modified lognormal distribution. *Entropy* 24(10):1362
- Hristopulos D, Pavlides A, Agou V, Gkafa P (2021) Stochastic local interaction model: an alternative to kriging for massive datasets. *Math Geosci* 53:1907–1949
- Islam Z, Ranganathan M, Bagyaraj M, Singh SK, Gautam SK (2022) Multi-decadal groundwater variability analysis using geostatistical method for groundwater sustainability. *Environ Dev Sustain* 24(3):3146–3164
- Keegan-Treloar R, Werner AD, Irvine DJ, Banks EW (2021) Application of indicator kriging to hydraulic head data to test alternative conceptual models for spring source aquifers. *J Hydrol* 601:126808
- Kitanidis PK (1997) *Introduction to geostatistics: applications in hydrogeology*. Cambridge University Press, Cambridge, UK
- Kitanidis PK, Vomvoris EG (1983) A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resour Res* 19(3):677–690
- Kovitz J, Christakos G (2004) Spatial statistics of clustered data. *Stoch Env Res Risk A* 18(3):147–166
- Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. *J Chem Metall Miner Soc S Afr* 52:119–139
- Manzione RL, Castrignano A (2019) A geostatistical approach for multi-source data fusion to predict water table depth. *Sci Total Environ* 696:133763
- Matheron G (1962) *Traité de géostatistique appliquée, tome 1 [Treatise on applied geostatistics, vol 1]*. Technip, Paris
- Mesić Kiš I (2016) Comparison of ordinary and universal kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac field. *Min Geol Pet Eng Bull* 31(2):41–58
- Olea R (2006) A six-step practical approach to semivariogram modeling. *Stoch Env Res Risk A* 20(5):307–318
- Olea RA (2012) *Geostatistics for engineers and earth scientists*. Springer, New York
- Parasyris A, Spanoudaki K, Varouchakis EA, Kampanis NA (2021) A decision support tool for optimising groundwater-level monitoring networks using an adaptive genetic algorithm. *J Hydroinformatics* 23(5):1066–1082
- Pavlides A, Hristopulos DT, Roumpos C, Agioutantis Z (2015) Spatial modeling of lignite energy reserves for exploitation planning and quality control. *Energy* 93:1906–1917
- Pavlides A, Agou VD, Hristopulos DT (2022) Non-parametric kernel-based estimation and simulation of precipitation amount. *J Hydrol* 612:127988
- Schabenberger O, Gotway CA (2004) *Statistical methods for spatial data analysis*. CRC, Boca Raton, FL
- Schrader A, Winde F (2015) Unearthing a hidden treasure: 60 years of karst research in the Far West Rand, South Africa. *S Afr J Sci* 111(5–6):1–7
- Tapoglou E, Karatzas GP, Trichakis IC, Varouchakis EA (2014) A spatio-temporal hybrid neural network-kriging model for groundwater level simulation. *J Hydrol* 519:3193–3203
- Tatas Chu H-J, Burbey TJ (2022) Estimating future (next-month's) spatial groundwater response from current regional pumping and precipitation rates. *J Hydrol* 604:127160
- Theodoridou P, Varouchakis E, Karatzas G (2017) Spatial analysis of groundwater levels using fuzzy logic and geostatistical tools. *J Hydrol* 555:242–252



- Varouchakis EA, Hristopulos D (2013a) Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environ Monit Assess* 185(1):1–19
- Varouchakis EA, Hristopulos DT (2013b) 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
- Varouchakis E, Hristopulos D, Karatzas G (2012) Improving kriging of groundwater level data using nonlinear normalizing transformations: a field application. *Hydrol Sci J* 57(7):1404–1419
- Varouchakis EA, Corzo GA, Karatzas GP, Kotsopoulou A (2018) Spatio-temporal analysis of annual rainfall in Crete, Greece. *Acta Geophys* 66(3):319–328
- Wackernagel H (2003) *Multivariate geostatistics: an introduction with applications*, 3rd edn. Springer, Berlin
- Yamamoto JK (2005) Correcting the smoothing effect of ordinary kriging estimates. *Math Geol* 37(1):69–94
- Zirakbash T, Admiraal R, Boronina A, Anda M, Bahri PA (2020) Assessing interpolation methods for accuracy of design groundwater levels for civil projects. *J Hydrol Eng* 25(9):04020042

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.