**ORIGINAL PAPER**

# Predictive modeling and comparative evaluation of geostatistical models for geochemical exploration through stream sediments

Muhammad Ahsan Mahboob[1,2] · Turgay Celik[3] · Bekir Genc[1]

## Abstract
Geochemical exploration of stream sediments is an important step for the identification of areas of interest with potential mineralization, particularly in the early stages of mineral exploration. A total of 407 sample points for 15 different geochemical traces were collected from Central Wales and classified into two groups: a training group consisting of 285 samples and a testing group consisting of 122 samples. Geospatial characterization of each parameter at stream level was performed using two different prediction models; the inverse distance weighting and the geostatistical kriging. Several variations of the IDW model was applied based on the power function and the number of the sample points, and the best one selected based on "root mean square prediction error" statistic. The same statistic was also used in the best-fitted semivariogram models including the circular, spherical, exponential and Gaussian for each geochemical parameter in Kriging. Finally, the mineral prospectivity map of the area was developed based on the geochemical accumulation index (GAI) using multivariate overlay analysis. The experimental results show that there is no single method that can be used independently to predict the spatial distribution of geochemical elements in streams. Instead, a combinatory approach of IDW and kriging is advised in order to generate more accurate predictions. The mineral prospectivity map based on GAI showed that most of the mineral-enriched streams were found in northwest, northeast, and south part of the study area which was also confirmed with the existing mining activities in the region.

**Keywords** Geochemical mineralization prediction · Geostatistics · Stream sediments · Geographic information systems · Spatial modeling · Mineral exploration

## Introduction

Stream sediments usually result from the erosion and transportation of soil and rock remnant, and other materials present upstream of the sampling locations within a basin (Merritt et al. 2003). Thus, stream sediments are considered to be representative of the geochemistry of the rocks present in the upstream drainage basin. Usually, stream sediments are the primary sampling source for geochemical exploration, primarily in areas where distinct drainage systems exist due to the local topography (Fletcher 1997; Salminen and Gregorauskien 2000; Zuluaga et al. 2017). Geochemical exploration sampling from stream sediments is a suitable technique for identification of areas of interest with significant mineralization content, particularly in the early stages of mining for undiscovered minerals (Arndt et al. 2017; Carranza 2017; Cheng 2007, 2012; de Mulder et al. 2016; Khalajmasoumi et al. 2017; Nieto et al. 2014). For the best results, it is always preferred to get as many samples as possible to work with a dense and large sampling dataset. However, high-density sampling of stream sediments is challenging due to the cost involved in getting geographical accessibility. Therefore, a limited number of samples are usually collected from the field and after analyzing in the lab used for estimation for larger areas using geostatistical interpolation models (Ottesen et al. 1989).

Geostatistical models have been extensively used as a powerful tool for providing accurate estimates of any spatial/

✉ Bekir Genc
   bekir.genc@wits.ac.za

1   School of Mining Engineering, University of the Witwatersrand, Johannesburg, South Africa

2   Sibanye-Stillwater Digital Mining Laboratory (DigiMine), Wits Mining Institute (WMI), University of the Witwatersrand, Johannesburg, South Africa

3   School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa

temporal phenomena at unsampled locations using sampled data along with a quantification of the related uncertainty (Cai et al. 2018; Dagdelen and Vega 1997; Goovaerts 2000; Liu et al. 2006, 2018; Lu and Wong 2008; Moral 2010; Piel et al. 2013; Ssempiira et al. 2017). Several spatial interpolation methods, including geostatistics, have been developed in the past to better predict spatial distributions based on limited samples. However, the accuracy of the interpolated surface varied widely between models (Robinson and Metternicht 2006). In general spatial interpolation methods can be divided into three main categories, i.e., with respect to the number of points used for prediction (global/local), with respect to the surface smoothness (exact/approximate), and with respect to the error estimation (deterministic/stochastic) (Li and Heap 2014).

The most common interpolation methods used for geospatial characterization of geochemical data from streams are Nearest Neighbor (NN) (Li and Heap 2014; Yang et al. 2004), Triangular Irregular Network (TIN) (Li and Heap 2014; Wu et al. 2011; Yang et al. 2004), Inverse Distance Weighting (IDW) (Li and Heap 2014; Lu and Wong 2008), Radial Based Functions (RBF) (Ding et al. 2017) and Kriging (Kleinschmidt et al. 2000; Panahi et al. 2004; Wu et al. 2011). In general, kriging is the most applied, tested, and validated geostatistical model in the field of spatial interpolation. The most significant advantage of this model over several other spatial and statistical models is that kriging considers the spatial correlation of the data for the predictions. Several types of kriging have evolved over the time, but the most commonly used are ordinary kriging (OK), simple kriging (SK), universal kriging (UK), indicator kriging (IK), and co-kriging (Panahi et al. 2004). Among all these techniques, the predictions are highly dependent not only on the location of the data but also on the accurate development of the semivariogram model. But the development of the required semivariogram model is time taking and subjective. In addition to that, the transformation of the data may also be required in non-stationary conditions and directional trends needed to be considered (Nieto and Toffait 2007). Hence different geospatial characterization methods have been used in the past to characterize geochemical mineralization which tends to overextend predictions to areas not really influenced by the original stream mineralization condition. As it is not required that a stream sediment sample ($S_1$) be collected at a particular location ($L_1$) on a stream which represents the geochemical value at $L_1$, it is quite possible that it is a representative of the upstream location $L_2$ or $L_3$, i.e., the potential source of the mineralized zone (Fig. 1).

Therefore this research focuses on the development of a new approach for geospatial characterization based on a continuous geochemical prediction surface up to the extent of sampling streams. Hence geochemically spatial predictions of continuous streams will provide improved estimations in streams considering the source of the mineralization.

Geochemical data are usually geospatial data as it can be expressed as $x$, $y$, and $z$, where $x$ and $y$ are location coordinates (latitude, longitude or easting, northing) and $z$ represents the recorded value (i.e., elements concentration) at those coordinates. Geochemical data are usually stored in a spatial format as point data with location and value, hence can be processed in geoinformatics (Zuo et al. 2016). With the advancement in computer technology, several geographic information system (GIS)-based numerical codes are available with a strong component of geostatistics embedded in them. Surface interpolations through geostatistical modelling in GIS are very powerful for estimating geochemical values (Johnston et al. 2012; Li and Heap 2014). This research study has also incorporated a very strong GIS numerical package ESRI's ArcGIS Geostatistical Analyst, which is equipped with advanced spatial geostatistical predictive models for predictive modelling. The two widely used interpolation methods, including the geostatistical (IDW and Kriging), were selected and applied to test their efficiency for geochemical characterization and prediction along the unsampled streams. Also these, models have been extensively used in the past, but none of them has been applied and tested for the prediction of geochemical values at individual stream levels, which is a new proposed methodology of this research study. Furthermore, the geochemical accumulation index (GAI) has been developed in this research to identify the mineral enriched streams and for identification of potential mineral prospects in the study area.

## Materials and methods

### Study area

The study area for this research is located in Central Wales of Great Britain (Fig. 2) and has a prolonged history of gold mining, with the main center of activity being the "Dolgellau Gold Belt" where the Clogau and Gwynfynydd mines were very active gold mines, particularly towards South, in the Welsh Basin, where gold has been mined since the Roman times. Besides gold, several other secondary products such as lead, copper, zinc, iron, nickel sulfides and others are present throughout the Wales area.

The dataset used in this study (WF/MR/93/013) was obtained from the British Geological Survey (BGS) (Brown 1993). The stream sediment baseline geochemistry data was estimated from samples collected across the central Wales region by BGS as shown in Fig. 2. Sampling was based on the collection of heavy minerals accumulated from the first- and second-order streams. Active stream sediment was moved through a 2-mm sieve, collected in a wooden pan (about 3–4 kg) and condensed by panning to about 60 g. This process was repeated using additional sediment from the same site and the two concentrates combined together were inspected on-site
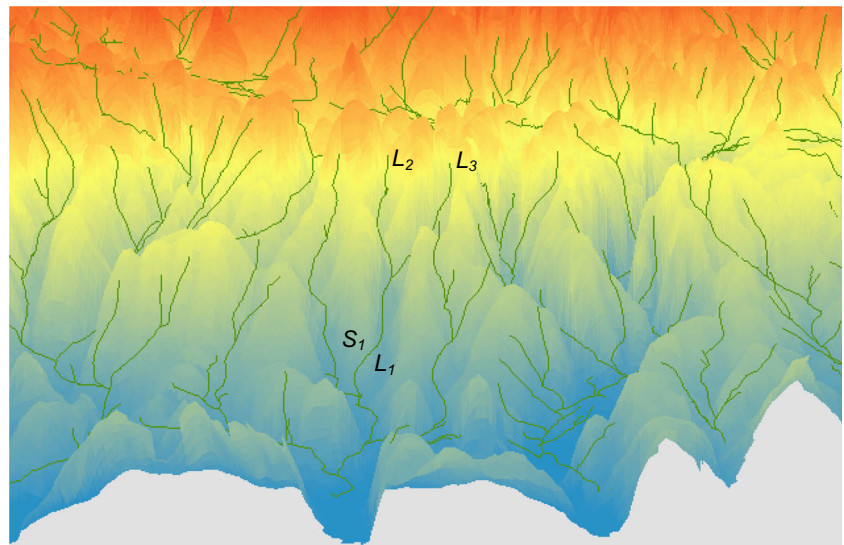
for heavy minerals and collected in a Kraft bag. A total of 407 samples were collected and analyzed for 15 different elements, including titanium (Ti), manganese (Mn), iron (Fe), vanadium (V), nickel (Ni), copper (Cu), zinc (Zn), zirconium (Zr), tin (Sn), antimony (Sb), barium (Ba), cerium (Ce), lead (Pb), arsenic (As), and gold (Au). Gold was estimated on 60 g of the sample by atomic absorption spectrometry after grinding and dissolution in aqua regia (a yellow-orange fuming liquid), with a lower confidence limit of detection as 10 parts per billion (ppb) Au. All other elements (Ti, Mn, Fe, V, Ni,
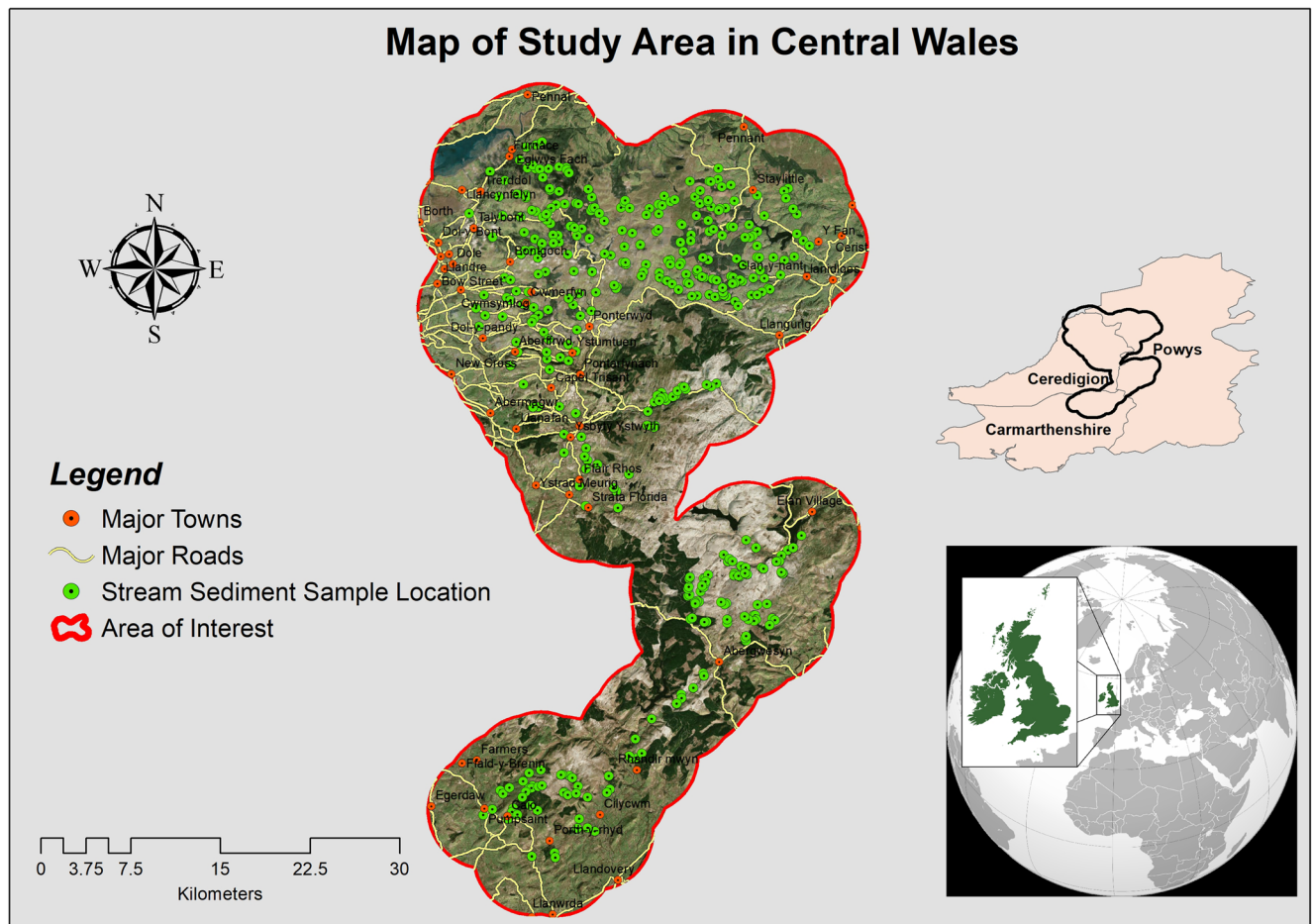


**Fig. 2** Study area map highlighting the location of major towns, roads, and stream sediment sample points

Cu, Zn, Zr, Sn, Sb, Ba, Ce, Pb, and As) were determined by X-ray fluorescence analysis of milled sub-samples in parts per million (ppm). The descriptive statistic of the 15 elements is given in Table 1.

From Table 1, it can be observed that the distributions of most geochemical parameters are positively skewed and their frequency diagrams do not follow a normal distribution (Bai and Ng 2005; Beedles and Simkowitz 1978). Coefficient of variation (CV) values < 10 and > 90% indicate low and high variability, respectively, in the data (Reed et al. 2002), and in this dataset, Ti and Au have the lowest and the highest variability, respectively. It can be observed through summary statistics that the data need to be transformed to a normal distribution and as all the variables have variability so spatial and geostatistical modelling can be applied for further predictions (Komnitsas et al. 2010). Further, the correlation co-efficient (Pearson coefficient) matrix, as shown in Fig. 3 was developed to highlight the statistical relationship between geochemical variables (Facchinelli et al. 2001). The correlation coefficient matrix reveals that the correlation coefficient value $r$ ranges from negative 0.66 (between Ti and Pb) to positive 0.69 (between Fe and V). The $r$ values are statistically significant at a 0.05 (5%) confidence level.

## Directional distributional trend of geochemical data

Usually, the standard deviational ellipse (SDE) is used to capture the spatial trend for a set of measured observations (Lefever 1926). The SDE is based on standard deviations computed on spatial coordinates ($x$, $y$) of the observations and has been applied to describe bivariate features. Most commonly, it is applied to assess the geographical distribution trend of the features concerned by analyzing both of their dispersion and direction. For a given set of $n$ samples $\mathcal{M} = \{(x_i, y_i, z_i)\}_{i=1}^n$, the corresponding SDE is defined by three parameters: spatial mean (or average spatial location) ($\overline{x}, \overline{y}$), spatial dispersion (or concentration) ($\sigma_x$, $\sigma_y$), and orientation of the geochemical data $\theta$. In addition to the traditional spatial mean center (gravity on the distribution), weighted spatial mean or median of the data could also be used as other options (Wang et al. 2015). The average spatial location of the SDE is computed using the following equation:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i, \overline{y} = \frac{1}{n}\sum_{i=1}^n y_i \tag{1}$$

the average spatial location is subtracted from each sample to center the observed samples at the origin, i.e.,

$$\widetilde{x}_i = x_i - \overline{x}, \widetilde{y}_i = y_i - \overline{y} \tag{2}$$

the centered samples are then used to compute the orientation of the SDE as follows

$$\theta = \arctan \frac{A + \sqrt{A^2 + 4B^2}}{2B} \tag{3}$$

where

$$A = \left(\sum_{i=1}^n \widetilde{x}_i^2 - \sum_{i=1}^n \widetilde{y}_i^2\right), \quad B = \sum_{i=1}^n \widetilde{x}_i \widetilde{y}_i$$

**Table 1** Summary statistics of geochemical elements in the study area ($N = 407$)

| Elements | Min | Max | Mean | Variation coefficient | Skewness (Pearson) |
|---|---|---|---|---|---|
| Ti | 730.0 | 7050.0 | 5247.0 | 19.6 | −2.0 |
| Mn | 10.0 | 9210.0 | 825.3 | 100.3 | 5.9 |
| Fe | 9400.0 | 76,100.0 | 46,805.7 | 25 | −0.8 |
| V | 0.0 | 360.0 | 102.7 | 36.3 | 0.0 |
| Ni | 7.0 | 656.0 | 47.9 | 107.6 | 7.9 |
| Cu | 1.0 | 19,917.0 | 549.4 | 414.5 | 5.4 |
| Zn | 33.0 | 17,454.0 | 777.2 | 309.4 | 4.7 |
| Zr | 0.0 | 304.0 | 175.6 | 22.4 | −1.2 |
| Sn | 0.0 | 260.0 | 10.9 | 243.4 | 5.6 |
| Sb | 0.0 | 693.0 | 14.0 | 414 | 7.1 |
| Ba | 62.0 | 2907.0 | 457.8 | 38.7 | 7.1 |
| Ce | 10.0 | 100,073.0 | 12,516.1 | 133 | 2.4 |
| Pb | 5.0 | 13,397.0 | 1157.4 | 248.5 | 2.6 |
| As | 0.0 | 80.0 | 12.8 | 67.5 | 3.3 |
| Au | 0.0 | 8765.0 | 36.3 | 1258.8 | 17.6 |
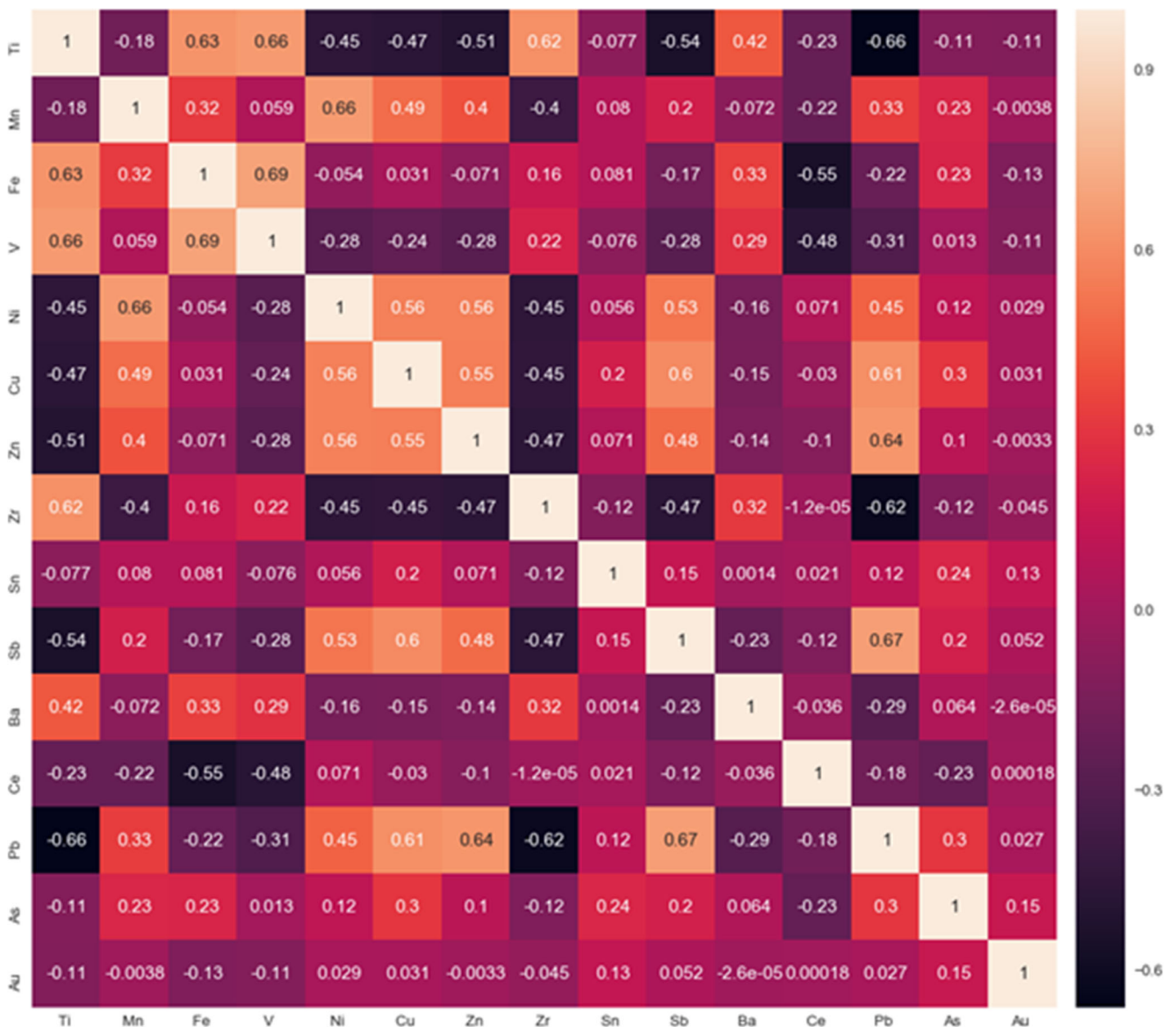
**Fig. 3** The correlation matrix between different geochemical variables highlighting the positive and negative associations

and finally, the spatial dispersion of the SDE is computed using the shifted and oriented samples based on

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_i\sin\theta + \widetilde{x}_i\cos\theta\right)^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_i\cos\theta - \widetilde{x}_i\sin\theta\right)^2} \tag{4}$$

A rule-of-thumb drawn from the Rayleigh distribution recommends that a first, second, and third standard deviational ellipse will cover approximately 63, 98, and 99.99% of the features (geochemical data) in two dimensions (x and y) (Wang et al. 2015). In this research, the directional distributional trend of all the 15 elements was assessed with one standard deviation, which covers 63% of the data in ArcGIS software.

## Streams delineation using digital elevation model

Global Digital Elevation Map (GDEM) derived from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) was used to extract the streams present in the study area. The GDEM was re-projected from Geographical Coordinate Systems to Projected Coordinates Systems for spatial referencing. The pre-processing of GDEM is always required and recommended before the extraction of any spatial information. Usually, a raw DEM often has a large number of "sinks" (or false depressions), i.e., single or multiple pixels with extreme low elevation and are completely enclosed by immediate higher elevation pixels (Jones 2002). The sinks, if present in the DEM, then must be removed using the fill operation available in any raster GIS packages. In this research study FILL operation of

ArcHydro model of ESRI's ArcGIS was applied as shown in Fig. 4. The FILL operation eliminates the sinks by either (a) raising the elevation on the sink cell to that of its lowest neighbor or (b) lowering the elevation of the lowest neighbor adjacent to the sink (Jones 2002).

After pre-processing of the GDEM, the next step is to determine the flow direction, which is the ability to control the direction of water from every pixel in the raster to its downslope neighbor. Several algorithms can be used to determine the direction of flow, including the Multi-Flow Direction (MFD), D-Infinity (DINF), and D8 (Płaczkowska et al. 2015). Among all these, the D8 (eight flow direction) method is the simplest and the most effective. D8 assigns the flow (in terms of coding) from individual pixel to one of its eight connecting cells, either adjacent or diagonal, in the direction with the steepest downward slope.

Once the flow direction raster is created, the next step is to calculate areas where water may accumulate as a result of rainfall, known as the Flow Accumulation raster. Mainly, the value to each cell in the resulting raster contains the sum of the amount of water that has fallen on all the raster cells upstream from it. The objective is to simulate the flow, or possible flow, of water to form creeks, streams, and/or rivers. A threshold value is required to determine the minimum number of cells that can generate the flow, i.e., in this study a threshold of 150 was set, which means that a minimum of 150 cells should contribute to generating a flow into a cell (Wei et al. 2016). The threshold value is selected from a range [100, 359] according to the number of connected pixels which fall on the stream to generate an adequate amount of flow. After highlighting the streams in the area, the next step is to convert them from raster to vector by using the Streams to Feature utility of ArcHydro. The mineral prediction at stream level will provide more realistic information and help to mark the mineral-enriched zones as compared with a classical square regional predicted surface.

## Spatial interpolation methods

### Inverse distance weighting (IDW)

IDW is a local, exact, and deterministic interpolation model, which means it predicts a value at unsampled locations using a subset of sampled values, and then it will be the same, i.e., the line passes through the points and gives no error assessment. Due to its simplicity and computationally non-intensiveness, several researchers, without having much knowledge in spatial statistics and geostatistics, will use IDW as a default method to produce a surface when data values exist only at sampled locations. This model assumes that data values that are nearby to one another are more similar than those that are at far distances (Wong 2016). To calculate a value for any unsampled location, IDW uses the measured values close to the unsampled location.

The IDW model predicts the unknown geochemical value $\widehat{g}^{(e)}(\mathbf{r}_u)$ of element $e$ at location $\mathbf{r}_u$, using the observed (or known) element values $g^{(e)}(\mathbf{r}_i)$ at sampled locations $\mathbf{r}_i$ as follows:

$$\widehat{g}(\mathbf{r}_u) = \sum_{i,\mathbf{r}_i \in \mathcal{N}_u^k} w_i g(\mathbf{r}_i), \tag{5}$$

where $w_i$ is the weight for the observed value $g^{(e)}(\mathbf{r}_i)$ for geochemical element $e$ and $\mathcal{N}_u^k$ is the set of $k$ nearest sample locations around the spatial location $\mathbf{r}_u$. In other words, the predicted value $\widehat{g}^{(e)}(\mathbf{r}_u)$ at $\mathbf{r}_u$ is a weighted sum of observed values $g^{(e)}(\mathbf{r}_i)$ with
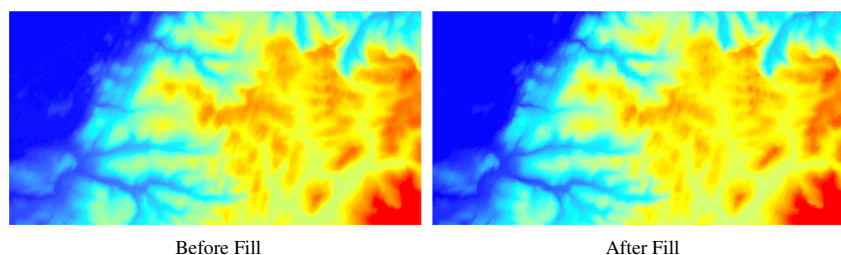
$$\sum_{i=1}^k w_i = 1$$

The weight $w_i$ is inversely proportional to the $\alpha^{\text{th}}$ power of Euclidean distance between $\mathbf{r}_u$ and $\mathbf{r}_i$ and defined as follows:

$$w_i = \frac{\|\mathbf{r}_i - \mathbf{r}_u\|^{-\alpha}}{\sum_{j=1}^k \|\mathbf{r}_j - \mathbf{r}_u\|^{-\alpha}}, \tag{6}$$

where $\|\mathbf{r}_i - \mathbf{r}_u\|$ is defined as Euclidean distance between $\mathbf{r}_u$ and $\mathbf{r}_i$.

The $\alpha$ parameter is quantified as a geometric form for the weight while other conditions are possible. This condition indicates that if $\alpha$ is more significant than 1, then the distance-decay influence will be more as compared with an increase in distance and vice versa. Therefore, a small $\alpha$ tend to give predicted values as averages of $g^{(e)}(\mathbf{r}_i)$ in the neighborhood, while large $\alpha$ assign more weights to the nearest points and gradually decrease for points far away. As a result, when $\alpha \to 0$



Fig. 4 The effect of FILL operation on the GDEM as on left side false depressions were filled after applying the operation

Before Fill

After Fill

$$\operatorname*{Lim}_{\alpha \to 0} \widehat{g}(\mathbf{r}_u) = \operatorname*{Lim}_{\alpha \to 0} \sum_{i=1}^{k} w_i g(\mathbf{r}_i);$$

$$= \operatorname*{Lim}_{\alpha \to 0} \sum_{i=1}^{k} \frac{\|\mathbf{r}_i - \mathbf{r}_u\|^{-\alpha}}{\sum_{j=1}^{k} \|\mathbf{r}_j - \mathbf{r}_u\|^{-\alpha}} g(\mathbf{r}_i);$$

$$= \operatorname*{Lim}_{\alpha \to 0} \sum_{i=1}^{k} \frac{1}{\sum_{j=1}^{k} 1} g(\mathbf{r}_i); \quad = \frac{1}{k} \sum_{i=1}^{k} g(\mathbf{r}_i),$$

$$\tag{7}$$

and $w_i \to 1/k$. Then, the predicted value is the average of all sampled values. Similarly when $\alpha \to \infty$, $\widehat{g}(\mathbf{r}_u) = g(\mathbf{r}_c)$ where $\mathbf{r}_c$ is the closest sample location.

Usually, the geoscientists and mineralogists use the power $\alpha = 2$, which is known as the inverse distance squared weighted model. In this research, three different IDW power values $\alpha = 1$, $\alpha = 2$, and $\alpha = 3$ had been tested to generate the geochemical prediction surfaces. There is no theoretical reasoning in the selection of a particular value over others, however, the influence of changing power should be examined by visualizing the output and observing the validation statistics. The measured values near to the unsampled location have more impact on the predicted value than those farther apart. IDW considers that each sampled point has a local impact that reduces with distance. So this model uses weights assigned to the sample values as it gives more weights to points nearby to the unsampled location, and the weights reduce as a function of distance; therefore, the name inverse distance weighting.

## Kriging

Kriging predicts the structure of spatial variability through a variogram and incorporates the spatial autocorrelation (Hattis et al. 2012). The kriging prediction is modeled as the summation of a global trend $\lambda$ (which is the general trend in the entire data) and a local stochastic variation $\varepsilon$ (Matheron 1963):

$$\widehat{g}(\mathbf{r}_i) = \lambda(\mathbf{r}_i) + \varepsilon(\mathbf{r}_i), \tag{8}$$

where $\mathbf{r}_i$ represents the spatial coordinates. Depending on the global trend $\lambda$, several types of Kriging models have been established. For example, simple kriging (SK) assumes $\lambda = 0$; ordinary kriging (OK) assumes $\lambda$ as an unknown constant mean; and universal kriging (UK) assumes $\lambda$ a general polynomial trend as $z = ax + by + c$, where $x$, $y$ are the variables for the latitude and longitude respectively and $z$ can be statistically analyzed from the past data. In this research study OK is used because it provides more realistic and reliable predictions based on mean squared errors; is an unbiased predictor for sparsely sampled regions (Cressie 1993); and minimizes the influence of spatial outliers (Triantafilis et al. 2001), if present in the data. The OK method predicts the $\widehat{g}(\mathbf{r}_u)$ at $\mathbf{r}_u$ using the weighted sum of the data as follows:

$$\widehat{g}(\mathbf{r}_u) = \sum_{i, \mathbf{r}_i \in \mathcal{N}_u^k} w_i \mathbf{g}_i(\mathbf{r}_i). \tag{9}$$

The choice of weight in Eq. (9) should be made in such a way that $w_i$ yields the lowest mean square estimation error (Cressie 1993). Besides this the selection of $w_i$ also depends on the type of semivariogram model as suggested by Deutsch and Journel (1992). The experimental semivariogram $\widehat{\gamma}(r)$ can be defined as the average square difference of the geochemical data values between the samples, parted by the lag vector $r$. If there are no explicit directional dependencies present among the data values, then Matheron method-of-moments estimator can be used to develop unidirectional experimental semivariogram as follows:

$$\widehat{\gamma}(r) = \frac{1}{2k(r)} \sum_{i=1}^{k(r)} \left\{ [g(\mathbf{r}_i) - g(\mathbf{r}_i + r)]^2 \right\} \tag{10}$$

where $k(r)$ is the number of sample pairs at lag distance $r$, (Deutsch and Journel 1992) after which $\widehat{\gamma}(r)$ is then fitted to the model function $\gamma_g(r)$. The weights $w_i$ for ordinary kriging can be given by the $(k+1) \times (k+1)$ linear systems of equations as follows:

$$\sum_{i, \mathbf{r}_i \in \mathcal{N}_u^k} w_i \gamma_g(\mathbf{r}_i, \mathbf{r}_j) + \mu = \gamma_g(\mathbf{r}_j, \mathbf{r}_u), \quad j = 1, \dots, k \tag{11}$$

$$\sum_{i, \mathbf{r}_i \in \mathcal{N}_u^k} w_i = 1 \tag{12}$$

where $k$ is the number of geochemical samples within the proximity of $\widehat{g}(\mathbf{r}_u)$; $\gamma_g(\mathbf{r}_i, \mathbf{r}_j)$ is the semivariogram between two geochemical samples $\mathbf{r}_i$ and $\mathbf{r}_j$; $\gamma_g(\mathbf{r}_j, \mathbf{r}_u)$ is the semivariogram between $\mathbf{r}_j$ and predicted $\mathbf{r}_u$; and $\mu$ is a linear external parameter called the Lagrange factor. The variance of OK prediction is given by Eq. (12), whereas the Lagrange factor compensates the uncertainty related to the mean value.

$$\sigma_{\mathrm{E}}^2 \widehat{g}(\mathbf{r}_u) = \sum_{i, \mathbf{r}_i \in \mathcal{N}_u^k} w_i \gamma_g(\mathbf{r}_i, \mathbf{r}_u) + \mu \tag{13}$$

## Semivariogram estimation

In literature, there are several types of semivariogram models (Olea 2006; Varouchakis and Hristopulos 2013), but the most commonly used are Gaussian and circular, and these are also tested in this study along with exponential and spherical. The general equation for all these semivariogram models are given as follows:

$$\gamma_g(\mathbf{r}) = \sigma_g^2 \left[ 1 - \exp\left( -\frac{|\mathbf{r}|}{\xi} \right) \right] \tag{14}$$

$$\gamma_g(\mathbf{r}) = \sigma_g^2 \left[ 1 - \exp\left( -\frac{|\mathbf{r}|^2}{\xi^2} \right) \right] \tag{15}$$

$$\gamma_g(\mathbf{r}) = \sigma_g^2 \left[ 1.5 \ |\mathbf{r}|/\xi - 0.5 \ (|\mathbf{r}|/\xi)^3 \right] \theta(\xi - |\mathbf{r}|) \tag{16}$$

where

$$\theta(\xi - |\mathbf{r}|) = \begin{cases} 0 & \text{if } \xi - |\mathbf{r}| < 0; \\ 1 & \text{if } \xi - |\mathbf{r}| \geq 0, \end{cases}$$

and

$$\gamma_g(\mathbf{r}) = \sigma_g^2 \left\{ 1 - \frac{2}{\pi} \cos^{-1}\left(\frac{\mathbf{r}}{\xi}\right) + \frac{2r}{\pi} \sqrt{1 - \frac{\mathbf{r}}{\xi^2}} \right\}. \tag{17}$$

in the above equations $\sigma_g^2$ is the variance; $|r|$ is the Euclidean norm of the lag vector $r$, and $\xi$ is the characteristics length. After the development of semivariogram models, the most appropriate one is selected by comparing the root mean square prediction error statistics between the observed and predicted geochemical values. Depending on the semivariogram model, the prediction error at location $S_0$ can be computed in terms of the standard deviation as follows:

$$\overline{E}(\mathbf{r}_u) = \sum_{i=1}^{k} w_i \gamma(\|\mathbf{r}_i - \mathbf{r}_u\|) \tag{18}$$

where $\|\mathbf{r}_i - \mathbf{r}_u\|$ is the distance among the locations $\mathbf{r}_i$ and $\mathbf{r}_u$.

### Sample size requirements for variogram computation

As per recommended in the literature, a minimum of 100–150 data points are required to achieve a steady variogram (Voltz and Webster 1990). This requirement is fulfilled in this research, with 407 geochemical points available for each element. Because the number of sample points was more than the minimum number as recommended in literature, so anisotropy (directional trend) can also be determined if present in the data. The ArcGIS Geostatistical Analyst can compute the optimum parameters, such as Major Axis (Range), Minor Axis (Range), and Angle of Rotation (Direction) to constitute the anisotropic effect.

### Potential mineral prospects

The potential mineral prospectivity was done by developing the geochemical accumulation index (GAI) using the multivariate overlay analysis to map the high, medium, and low-mineral enriched streams. Multivariate overlay analysis has been used by several research studies (Correia and Waitzberg 2003; Hou et al. 2017) to classify the regions in different classes of underlying physical phenomenon. All the predicted surfaces of both IDW and Kriging were classified into three main classes based on Jenks Natural Breaks (JNB) algorithm (Khamis et al. 2018). With JNB, classes were made for each geochemical predicted map based on features similarity and

relatively significant data value differences in classes. The standardized classes were combined through linear combination by adding them together to obtain the final high, medium, and low mineralized streams.

## Results and discussion

The directional distributions (standard deviational ellipse) were used to map the geographical trend of different geochemical elements, as shown in Fig. 5. The ellipses were developed using the first standard deviation (almost 63% of the data) as it is recommended by several research studies (Kent and Leitner 2007; Seidl et al. 2015) to avoid the effects of any potential spatial outlier if present in the data.

In general, all the elements are distributed north to south except the gold, which showed the directional trend in northeast to southwest. For a detailed analysis, these 15 geotechnical elements were grouped in five different classes of direction and geographical concentration, as shown in Fig. 5.

The results showed that nine geochemical elements (As, Ba, Ce, Fe, Mn, Ni, Ti, V, Zr) have north to south distribution, whereas Mn and As are concentrated more towards the north and south, respectively. The main regions of As mineralization may be due to metasedimentary rocks and form post-orogenic granites in the southeast area of the study. Cohen et al. (1999) also found similar results in Wales and associated the variability of As with the geological rocks located within the area. The three elements Pb, Sb, and Zn can be grouped because of the same directional trend, i.e., northwest to southeast with Zn as a bigger ellipse as it has more variability. The Sn also has north-south concentration but the largest deviational ellipse, which means it has the highest variability and 63% of the values distributed all over the study area. Cu has the most concentrated values, i.e., smaller standard deviational ellipse and is mainly distributed north-south but at west of the study area. The gold has an altogether different distribution, which is northeast to southwest, but 63% of the data values are geographically concentrated in the lower part of the study area.

### Streams delineation

Stream delineation was performed from the ASTER global digital elevation model with 30-m spatial resolution using the D8- algorithm. The flow direction raster shows that the majority of the flow in the basin was from north and northeast to southwest and west which can be due to the presence of the Irish Sea on the West of Wales. The flow accumulation raster shows the major flow lines (streams) in the area based on the directional raster.

**Fig. 5** Directional distribution through standard deviational ellipses of all the geochemical elements (N = 407)
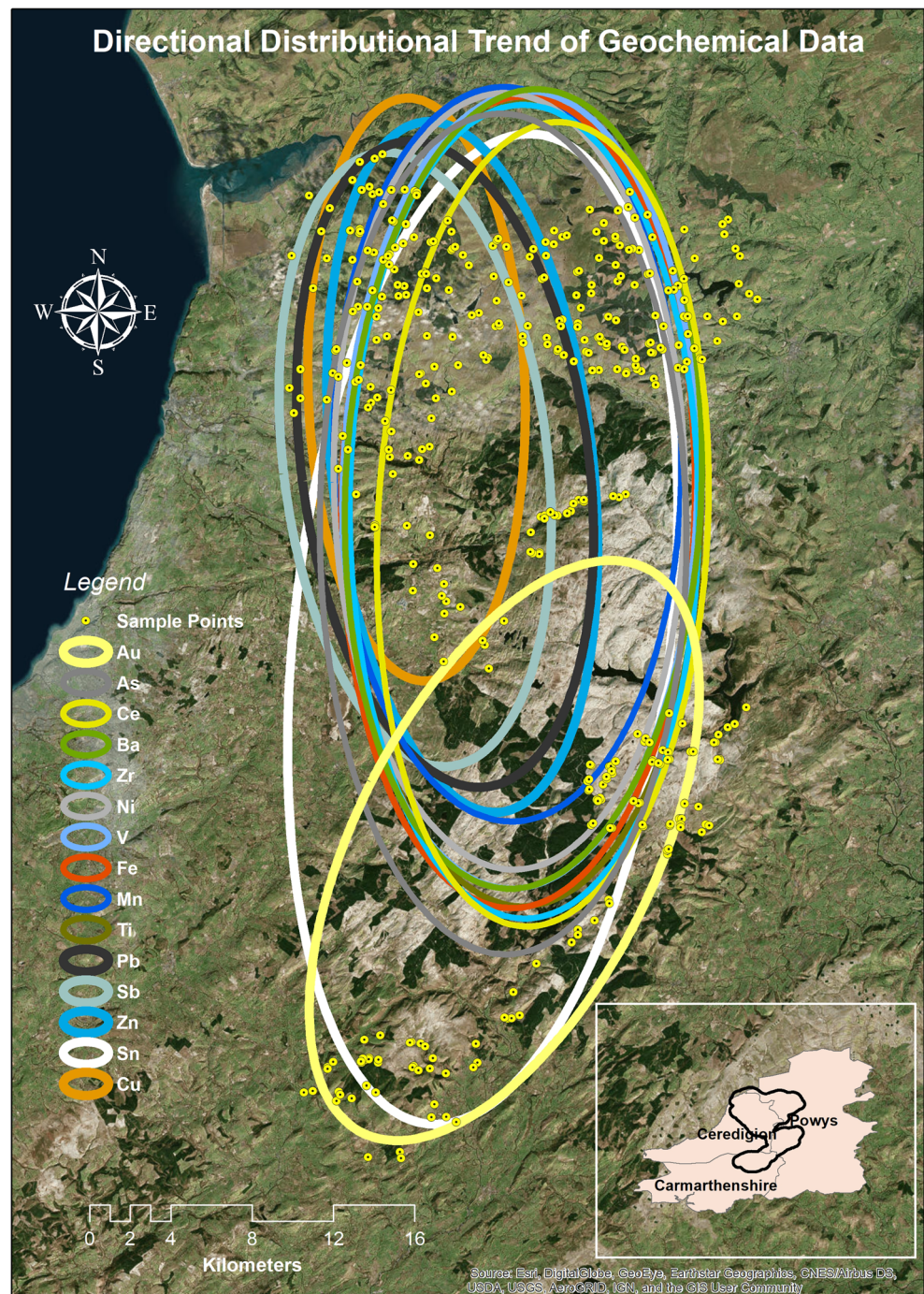


Finally, the streams were extracted using that flow accumulation raster with a threshold value of 150, as discussed in the methodology section of this paper.

Figure 6 showed that all the sample locations were correctly overlaid on the streams, which can be taken as the verification of the accuracy of extracted streams and the used threshold value. These streams were further converted into a vector data format, and a buffer of 100 m was applied to them, to be further used in the prediction models, as shown in Fig. 7.

## Criteria for comparison

The geochemical sample points were split into two groups named as the training and the test group with 70% and 30 of the points, respectively. The criterion to split data was that the sample points located in the homogenous geological group and on the same streams were selected and separated from the training data as test data.

Out of the total 407 geochemical sample points, 285 were used to train the model, and the remaining 122 points were

**Fig. 6** Extracted streams (solid
blue lines) from the digital
elevation model and overlaid on
the sample locations (solid green
dots)
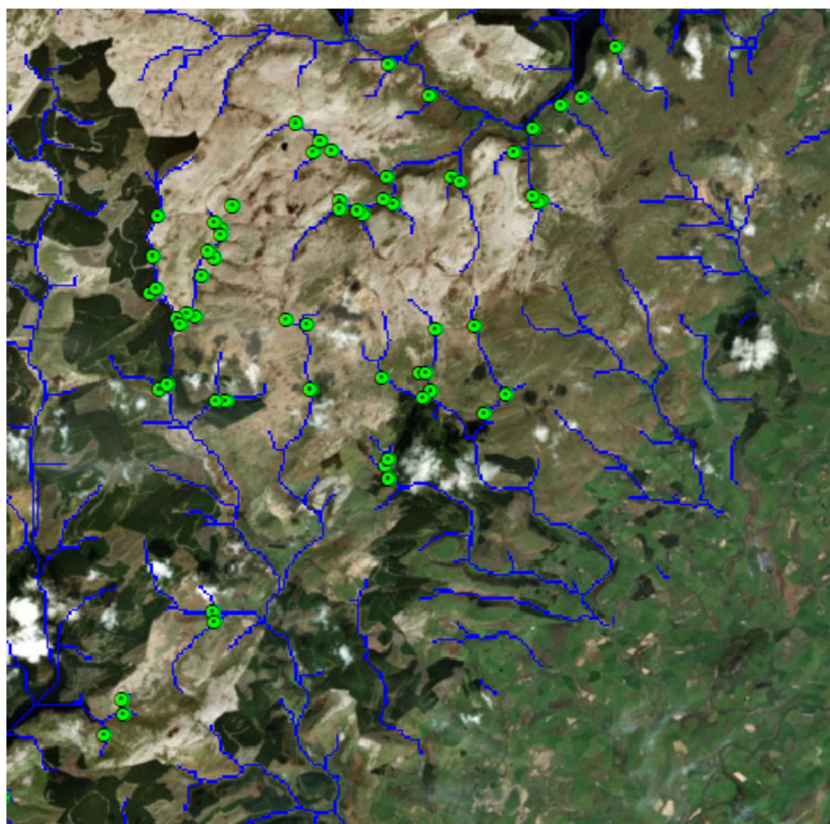


used to test the accuracy of the predicted surfaces. The root
means square prediction error (RMSPE) statistics were used to
select to assess the difference between predicted and the actual
geochemical values (Robinson and Metternicht 2006). The
formula for RMSPE is given as the following equation:

$$\text{RMSPE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\{Z(x_i) - \widehat{Z}(x_i)\right\}^2} \qquad (19)$$

where $\widehat{Z}(x_i)$ is the predicted geochemical value, $Z(x_i)$ is the
geochemical observed (known) value, and $N$ is the total num-
ber of samples. Ideally, the value of RMSPE should be zero if
there is no error in prediction; however, the low RMSPE sug-
gests less error in the predicted surface and vice-versa.

## Spatial interpolation and interpretation

The predicted/interpolated geochemical surfaces developed
by IDW were controlled and changed by varying the two
IDW parameters, i.e. the power function and the number of
sample points used in the prediction. Starting with the lower
power ($\alpha$) as 1 to a higher power of 3 were tested, whereas the
number of the nearest sample points were varied from 5 to 15
with increments of 5 at each step. The other parameters, like
sector type, which is the directional search for the inclusion of
several points, were kept fixed, i.e., four sectors with a 45°

offset, angle 0, and auto-calculation of semi-major and minor
axis.

For an initial prediction, the model was set with low power
($\alpha$) as 1 and a maximum of 5 neighbor points were selected.
The number of neighbor points was increased up to a maxi-
mum of 15 with the same power. The same procedure was
repeated for powers 2 and 3. The results showed that a single
power factor and number of neighbors could not be used for
all the geochemical variables; all the three power value and a
different number of neighbors produced best geochemical pre-
diction surfaces for different elements. The lowest root mean
square prediction error (RMSPE) statistics (Table 2) were
used to select the most appropriate combination model param-
eters for the geochemical elements.

To discuss the effects of power and the number of neigh-
bors Sn is taken as a case study, the prediction surfaces with
all the variation of powers and neighbors are shown in Fig. 8.

The comparison of Sn-predicted surface, as shown in
Fig. 8, revealed that areas with a low (near to zero) con-
centration remain almost the same in all the model types,
but the area of medium and high concentration varies with
different models. The change is quite significant at the
edges and in the middle, particularly the southern part of
the map. The maps of all the predicted streams geochem-
ical properties through the IDW models based on the low-
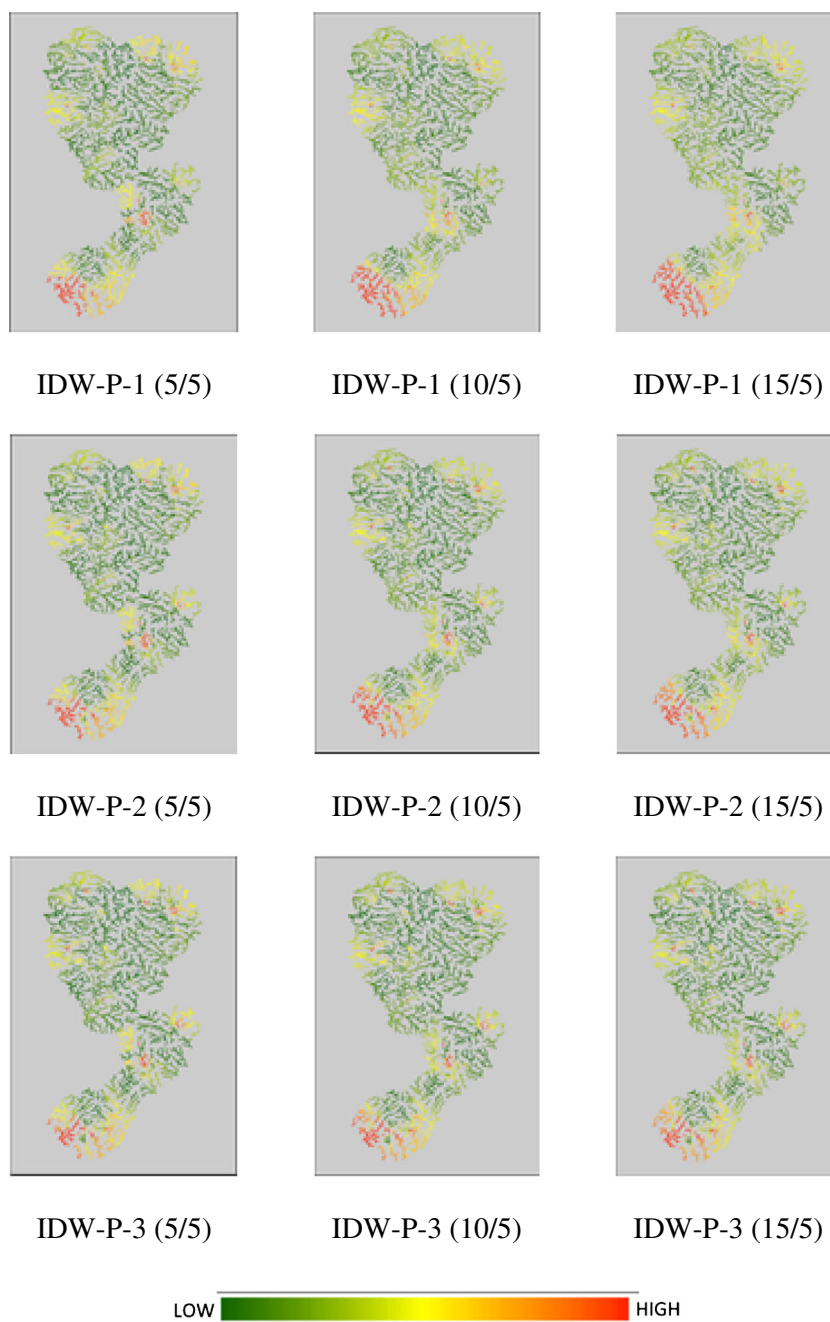est RMSPE are given in Fig. 9.

**Fig. 7** The extracted streams in
the study area



**Delineated Streams in the Study Area**

Legend

■ Streams

0   5   10   20
Kilometers

**Table 2**   RMSPE of all the geochemical elements against different IDW models

|     | IDW-P-1 (5/5) | IDW-P-1 (10/5) | IDW-P-1 (15/5) | IDW-P-2 (5/5) | IDW-P-2 (10/5) | IDW-P-2 (15/5) | IDW-P-3 (5/5) | IDW-P-3 (10/5) | IDW-P-3 (15/5) |
|-----|---------------|----------------|----------------|---------------|----------------|----------------|---------------|----------------|----------------|
| Ti  | 935.306       | 930.288        | 928.567        | 982.25        | 970.99         | 966.118        | 1030.16       | 1022.93        | 1020.6         |
| Mn  | 579.784       | 599.78         | 615.654        | 551.378       | 550.218        | 549.433        | 551.375       | 549.231        | 547.605        |
| Fe  | 10,176.4      | 10,103.139     | 10,131.154     | 10,801.5      | 10,658.5       | 10,602.4       | 11,411.1      | 11,327.8       | 11,301.3       |
| V   | 34.497        | 34.064         | 34.213         | 36.225        | 35.652         | 35.54          | 38.311        | 37.98          | 37.918         |
| Ni  | 41.582        | 42.233         | 42.779         | 38.871        | 38.621         | 38.51          | 38.632        | 38.43          | 38.328         |
| Cu  | 1624.36       | 1650.882       | 1669.518       | 1698.91       | 1683.29        | 1674.35        | 1799.99       | 1795.61        | 1793.09        |
| Zn  | 2047.331      | 2063.756       | 2079.226       | 2130.86       | 2113.68        | 2105.43        | 2278.01       | 2269.98        | 2266.43        |
| Zr  | 32.354        | 32.309         | 32.658         | 34.193        | 33.917         | 33.838         | 35.523        | 35.394         | 35.349         |
| Sn  | 26.992        | 26.824         | 26.695         | 28.873        | 28.634         | 28.453         | 29.796        | 29.704         | 29.643         |
| Sb  | 46.614        | 46.112         | 46.034         | 45.513        | 45.329         | 45.508         | 48.993        | 48.634         | 48.566         |
| Ba  | 186.463       | 182.52         | 181.299        | 201.156       | 198.719        | 197.45         | 213.379       | 212.64         | 212.31         |
| Ce  | 14,657.94     | 14,734.557     | 14,827.495     | 15,512.6      | 15,383.3       | 15,353.4       | 16,499.4      | 16,410.6       | 16,394.1       |
| Pb  | 2305.09       | 2322.459       | 2348.981       | 2324.85       | 2318.12        | 2320.98        | 2387.3        | 2381.34        | 2380.93        |
| As  | 7.599         | 7.616          | 7.654          | 7.681         | 7.669          | 7.677          | 7.88          | 7.867          | 7.865          |
| Au  | 525.482       | 513.607        | 509.136        | 585.032       | 580.134        | 578.308        | 619.582       | 618.648        | 618.407        |

**Fig. 8** Prediction maps of Tin (Sn) for different power and sample points used in the IDW model



IDW-P-1 (5/5)  IDW-P-1 (10/5)  IDW-P-1 (15/5)

IDW-P-2 (5/5)  IDW-P-2 (10/5)  IDW-P-2 (15/5)

IDW-P-3 (5/5)  IDW-P-3 (10/5)  IDW-P-3 (15/5)

LOW ▭ HIGH

The results of geochemical properties of stream sediments showed that Au was present in the lower streams of the study area towards the south and southwest, whereas no or very low gold concentration was found in the northern part of the study area. There were only traces of Gold present in the study area, and distinct anomalies in the area could be associated with the dominated alluvium (containing several significant placer deposits) or metasedimentary rocks (Cohen et al. 1999). As, Fe, Ti, V, and Zr had variable concentration and could be found throughout the streams present in the study area. The highest and above-average concentration of Ba was observed in the eastern and western streams to the north, respectively. The

high concentrations of Cu were mainly in the middle streams towards the north of the study area, whereas average values were found in the middle. The Ce was significant in the lower streams of the study area towards the south with a little higher amount in the northwest region of the study area. Mn showed less variability and is mainly concentrated in the upper streams more towards the north. Ni and Pb were found in high concentration near the eastern streams of the study area. Sb followed the same trend of concentration as Ni and Pb but in more streams. Sn was mainly dominated in the lower streams towards the south with medium and a few high quantities in the northwest and northeast of the study area. The highest
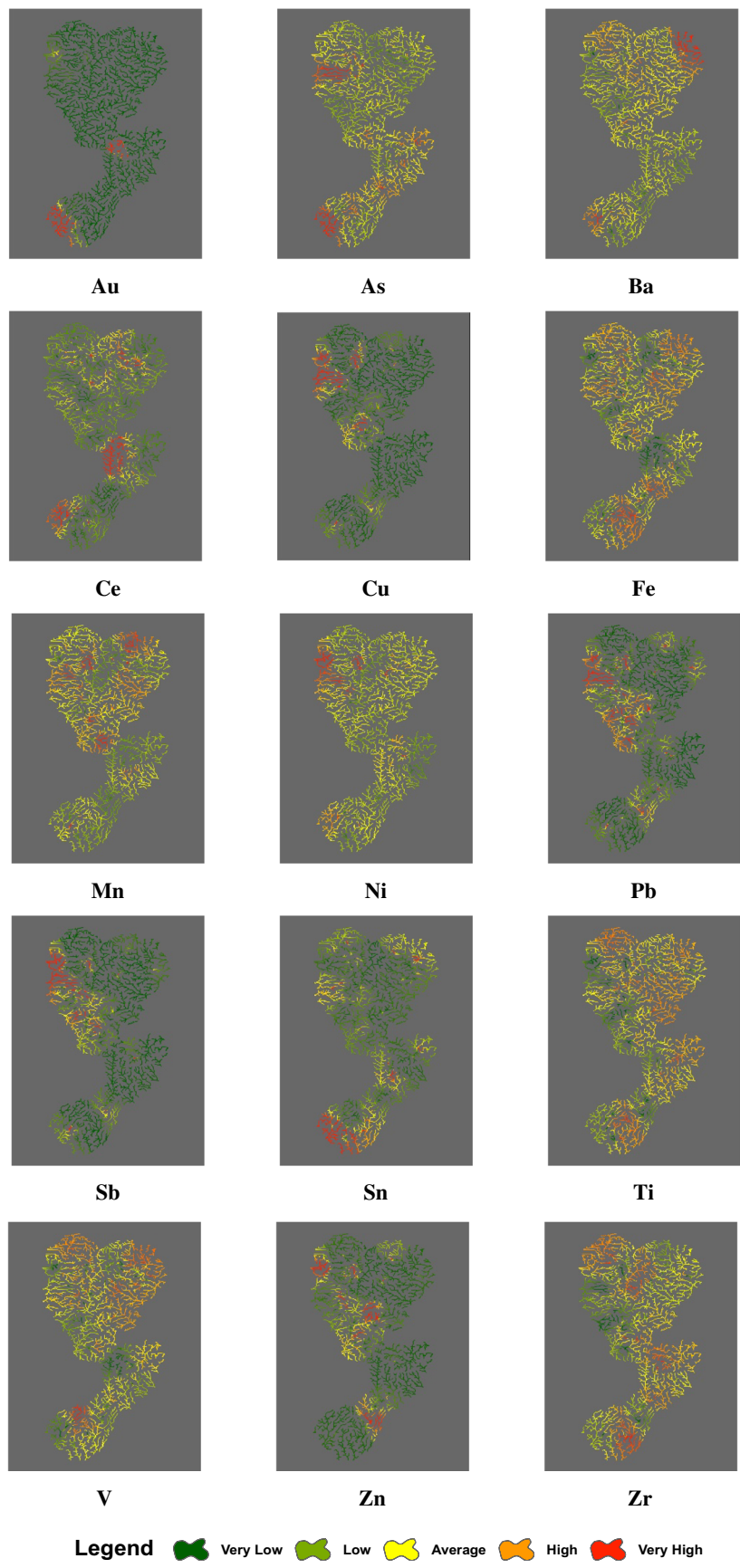
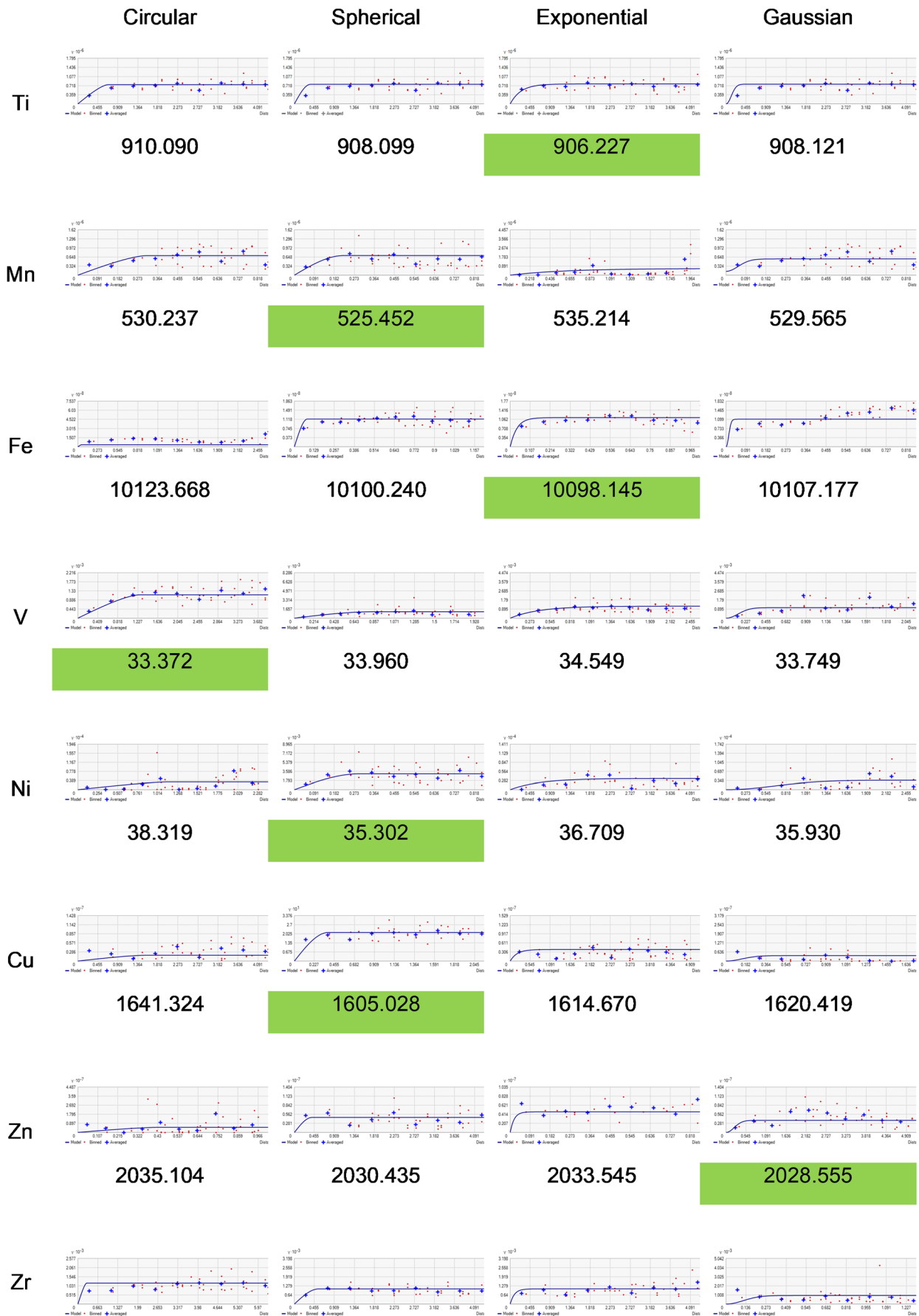**Fig. 9** Predicted stream surfaces of all the geochemical elements based on the lowest RMSPE of IDW model

|  | Circular | Spherical | Exponential | Gaussian |
|---|---|---|---|---|
| Ti | | | | |
| | 910.090 | 908.099 | 906.227 | 908.121 |
| Mn | | | | |
| | 530.237 | 525.452 | 535.214 | 529.565 |
| Fe | | | | |
| | 10123.668 | 10100.240 | 10098.145 | 10107.177 |
| V | | | | |
| | 33.372 | 33.960 | 34.549 | 33.749 |
| Ni | | | | |
| | 38.319 | 35.302 | 36.709 | 35.930 |
| Cu | | | | |
| | 1641.324 | 1605.028 | 1614.670 | 1620.419 |
| Zn | | | | |
| | 2035.104 | 2030.435 | 2033.545 | 2028.555 |
| Zr | | | | |



**Fig. 10** Semivariogram models of all the geochemical elements along with different RMSPE

🖄 Springer

| Sn | 26.346 | 25.286 | 28.331 | 27.197 |
| Sb | 49.526 | 45.621 | 43.955 | 46.515 |
| Ba | 178.687 | 182.180 | 179.187 | 183.357 |
| Ce | 14679.096 | 14632.843 | 14671.989 | 14778.979 |
| Pb | 2308.287 | 2311.019 | 2304.180 | 2316.875 |
| As | | | | |

**Fig. 10** continued.

concentration of Zn was observed in the northeastern and middle streams with average values in the lower streams of the study area. For kriging, interpolation is the first and the most important thing to develop an empirical semivariogram model and then fitting that model as structural analysis in Kriging; this process is usually known as variography. The four most commonly applied theoretical semivariogram models circular, spherical, exponential, and Gaussian were tested against the streams' sediment geochemical data. The semivariogram models were tried to fit the average of the binned data; also the directional variations were incorporated for micro-tuning of the models.

The selection of the most appropriate semivariogram model for a particular geochemical element was made based on the lowest yielded RMSPE by each model. The semivariogram of all the models, along with their RMSPE, is given in Fig. 10.

**Fig. 11** Prediction maps of Tin (Sn) for different semivariogram models used in kriging interpolation



**Circular**

**Spherical**

**Exponential**

**Gaussian**

LOW ———————————————— HIGH
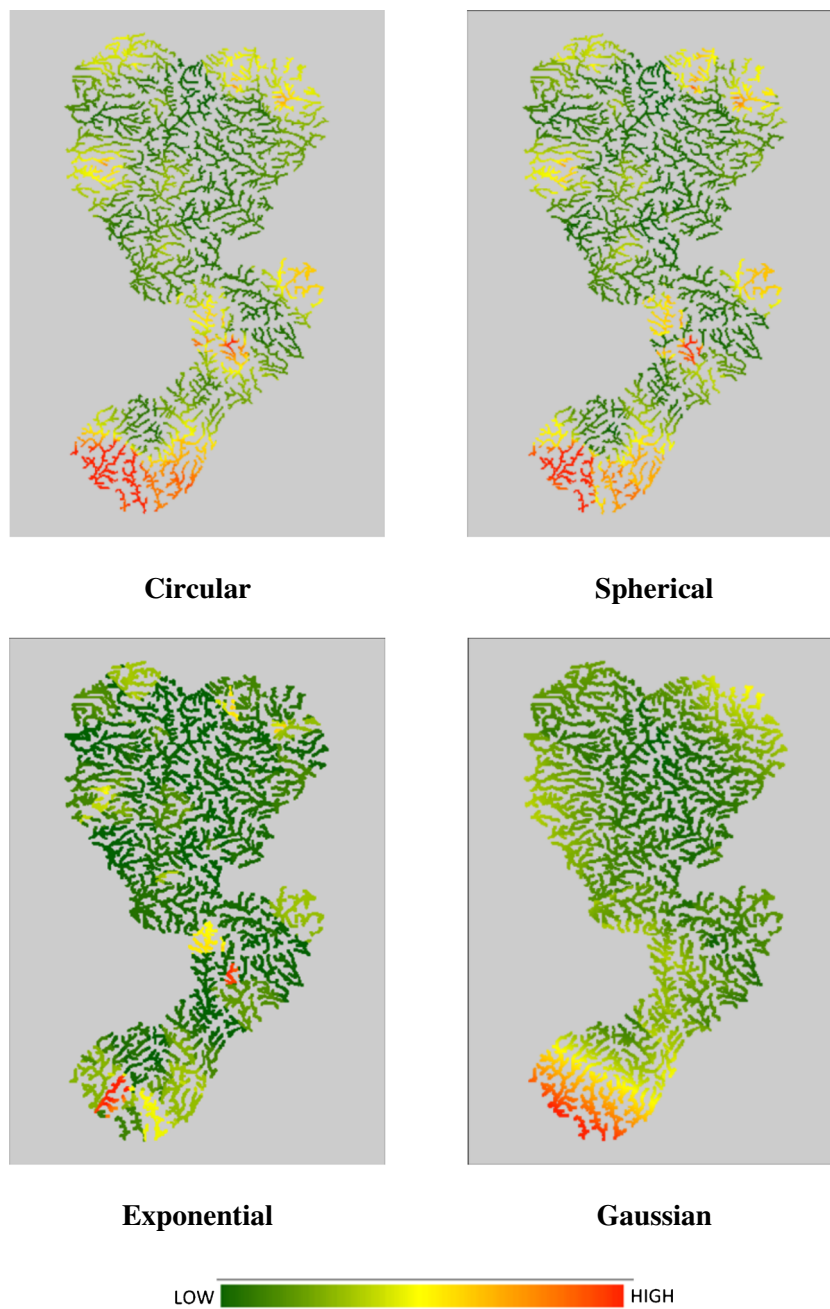
The analysis showed that out of the 15 streams' sediment geochemical elements, only two can be better predicted with the circular semivariogram model, which is V and Ba. The spherical semivariogram model can be used for Mn, Ni, Cu, Zr, Sn, Ce, As, and Au, whereas the exponential semivariogram model is suited best for Ti, Fe, Sb, and Pb. Only one geochemical element Zn can be best predicted by the Gaussian semivariogram model. To discuss the effects of the semivariogram model on stream sediments, Sn was taken as a case study, the prediction surfaces with all the circular, spherical, exponential, and Gaussian semivariogram models are shown in Fig. 11.

The comparison of Sn predicted surface through four semivariogram models, as given in Fig. 11, showed that the spherical model is more accurate than the others as it produced more continues geochemical values in streams. The same is supported by the semivariogram model of Sn, as mentioned in the Fig.10. The circular model also showed some continuity but is not as accurate as of the spherical. This continuity might be due to the circular semivariogram model, which shows a little variation in the start and then levels off. The exponential semivariogram model of Sn suggested abrupt changes in the geochemical values, and the same is highlighted in the
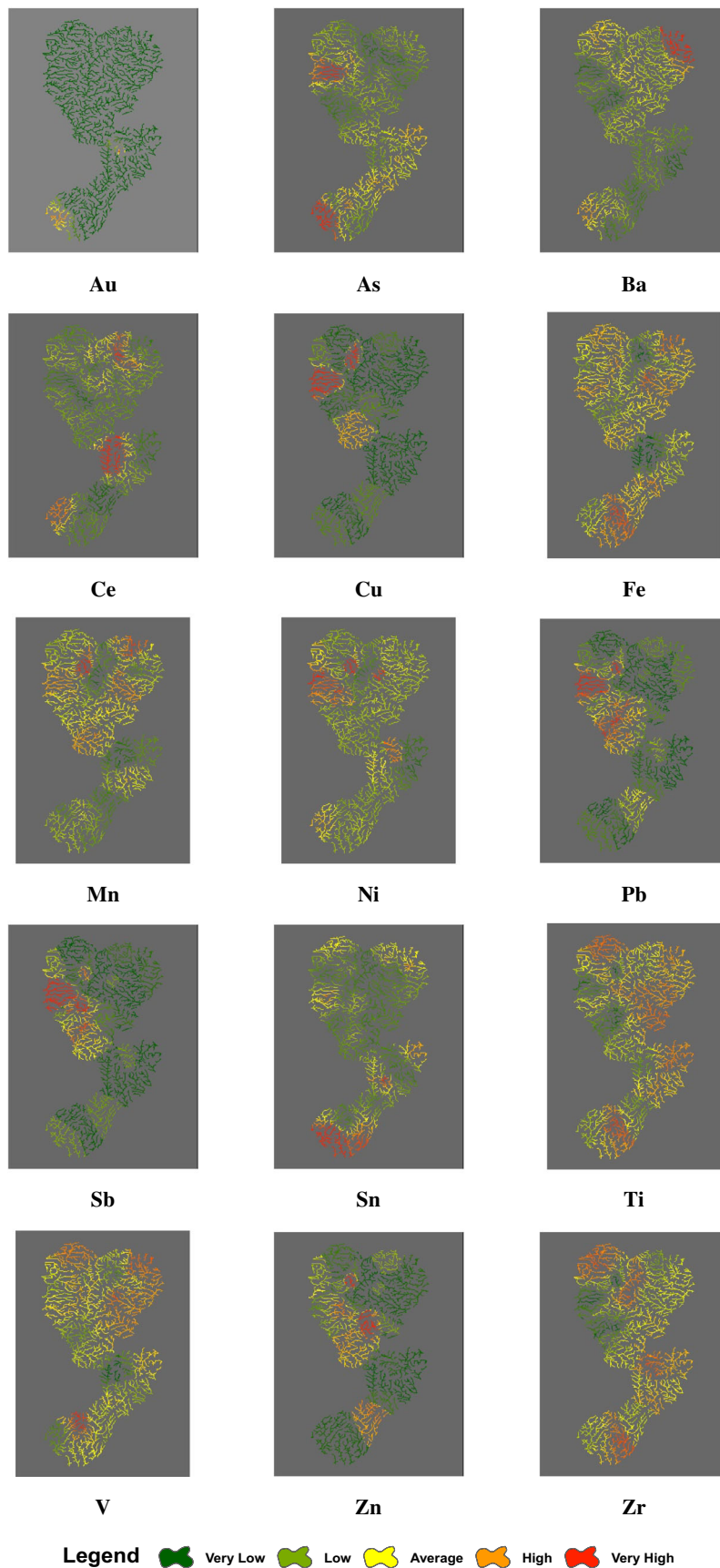
**Fig. 12** Predicted stream surfaces of all the geochemical elements based on the lowest RMSPE of semivariogram models using kriging
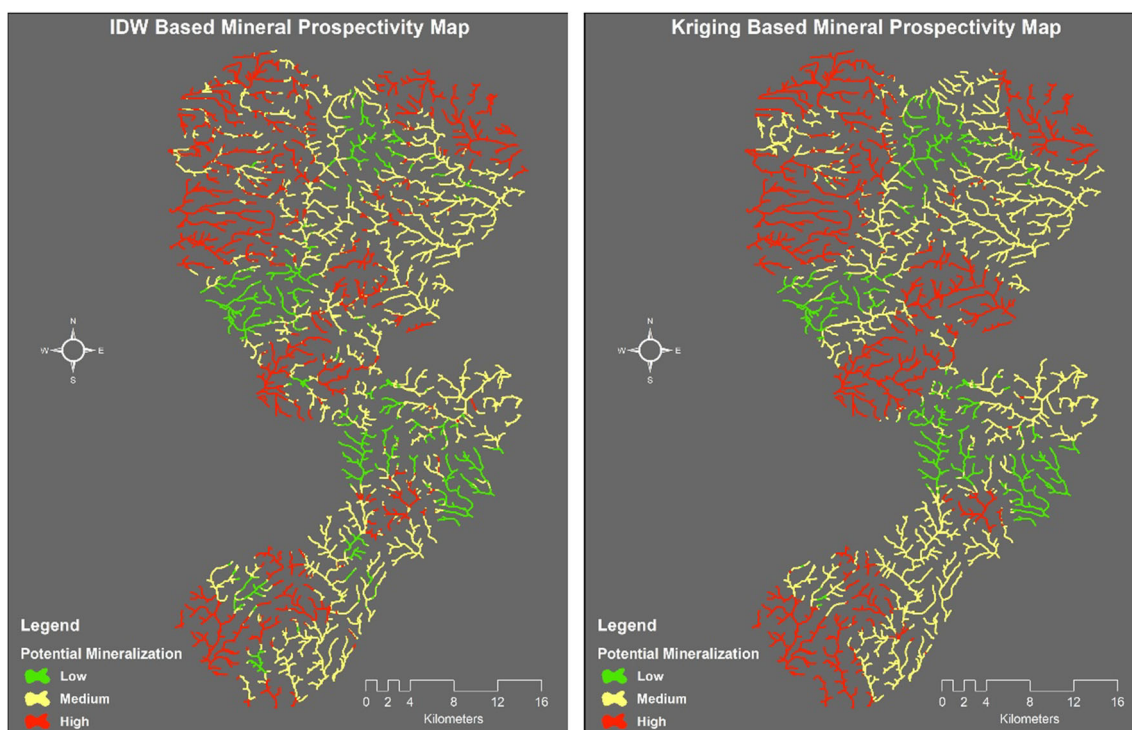
Fig. 13 Mineral-enriched streams extracted through geochemical accumulation index (GAI)

predicted surface. The Gaussian model showed some variability but is unable to depict appropriately because most of the geochemical sample points were not fitted through the model. The change is quite significant at the edges in the north and the middle, particularly the southern part of the map. The maps of all the predicted streams geochemical properties through the kriging model based on the lowest RMSPE of the best-fitted semivariogram model is given in Fig. 12. The trend of prediction surfaces by kriging is in line with the IDW results but significantly more realistic and continues.

## Potential mineral prospects

The multivariate overlay analysis for both IDW and kriging predicted surfaces were carried out in order to develop the geochemical accumulation index (GAI) of the study area, as shown in Fig. 13.

The higher values of GAI in a stream represent the higher chances of mineralization at the origin and surrounding regions and vice versa. The results showed that most of the mineral enriched streams were found in the northwest, northeast, and southern parts of the study area. By tracing these mineral-enriched streams, the potential mineral prospects can also be mapped accordingly and with less time and cost. The potential mineral prospects, as identified in this research area, were also confirmed with the existing mining activities in the regions, as

shown in Fig. 14. This showed that the methodology adopted in this research study could successfully be applied to identify the potential mineral prospects in a region using stream sedimentation.

## Comparison of prediction models

In order to choose the best prediction surfaces, the root means square prediction error (RMSPE) of each geochemical element for both IDW and Kriging models were compared (Table 3).

The highest and the lowest difference between IDW and Kriging model were found among Ce and As, respectively. For Ce and As, the IDW model with power as 1 and number of nearest sample points as 5 gives the better prediction surface with the lowest RMSPE, whereas as per kriging, the spherical model generates the lowest RMSPE which is also lower than the IDW model. Overall, the kriging model yields a low RMSPE because it incorporates the spatial correlation (autocorrelation) of the sample streams' geochemical values along with their locations. Several researchers also concluded that the kriging model is better for spatial predictions than the IDW model. For example, Shahbeik et al. (2014) applied IDW and kriging for mineral ore prediction, and it was concluded that the kriging performs much better, and the results were reliable. Similarly, another study conducted by Mueller et al. (2004) showed that kriging is more efficient in the prediction of soil fertility than IDW. Also, Goovaerts (2000) stated that

**Fig. 14** Existing mining activities overlaid on the GAI based mineral enriched streams
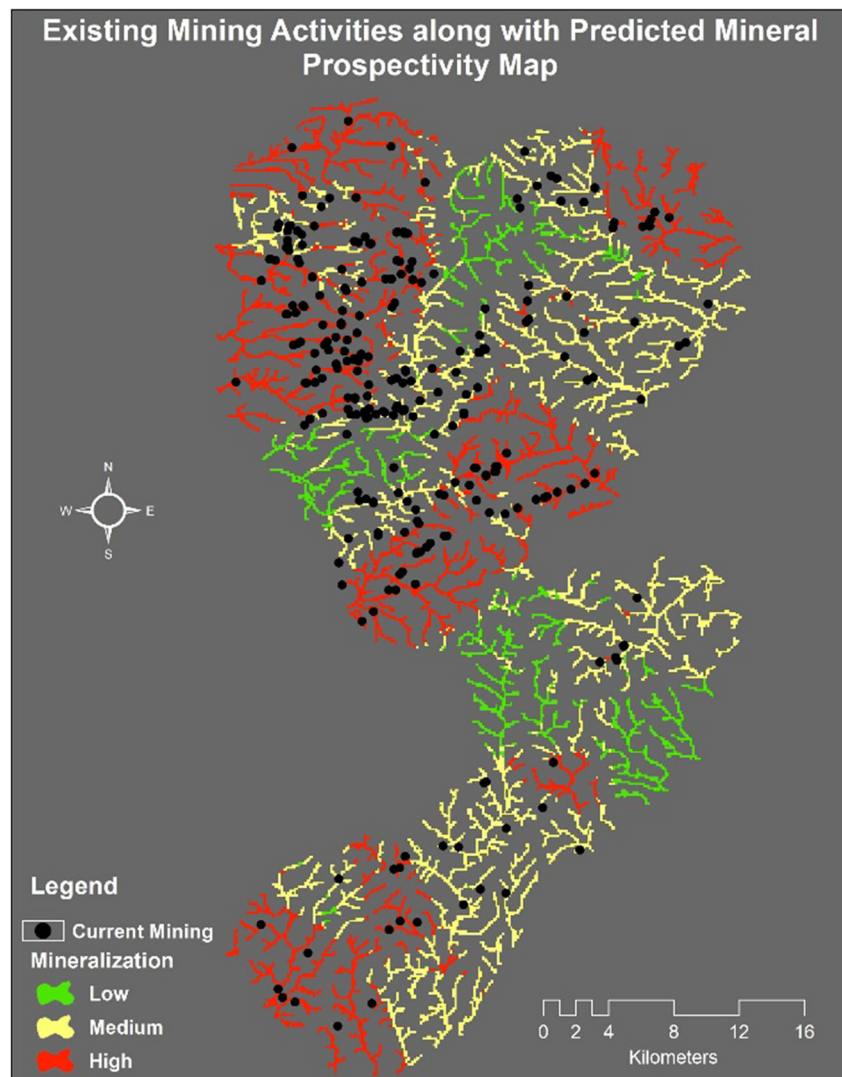


Kriging yields more accurate rainfall predictions with the lowest root mean square error.

## Conclusion

This study has shown a novel approach for prediction of geochemical properties of stream sediments using geostatistical interpolation models, including IDW and kriging. Instead of generating over-extended predicted surface models that most likely are outside the mineralization range from the stream samples, the new proposed geospatial characterizations were modelled strictly based on the geo-profile of each stream using digital elevation 3D models. The predicted geochemical value distribution at each stream showed an overall improved spatial characterization that can be used to accurately trace back the key minerals within each of the streams. The study showed that the two spatial prediction models used, IDW and kriging methods, should be used under a combinatory

approach in order to predict the spatial distribution of geochemical elements better. Overall, spatial realizations using IDW with a low power function tended to produce more accurate results probably due to the relatively high variability among the geochemical sample data (Robinson and Metternicht 2006) (Kravchenko and Bullock 1999). The kriging method also showed accurate results as well, specifically when using the spherical semivariogram model and then exponential models. The mineral prospectivity map based on the geochemical accumulation index (GAI) showed that most of the mineral-enriched streams were located in the northwest, northeast, and southern parts of Central Wales; the same was also confirmed with the existing mining activities in the region.

The statistics 'root mean square prediction error' (RMSPE) was used as the primary indicator to assess the estimation accuracy of both methods. The resulting models with the lower RMSPE values were considered for further analysis. One of the critical conclusions resulting from this study is that IDW

**Table 3** Semivariogram models of all the geochemical elements along with different RMSPE

| Geochemical element | IDW | Kriging | Difference |
|---|---|---|---|
| As | 7.599 | 7.474 | 0.125 |
| Au | 509.136 | 507.338 | 1.798 |
| Ba | 181.299 | 178.687 | 2.612 |
| Ce | 14,657.941 | 14,632.843 | 25.098 |
| Cu | 1624.36 | 1605.028 | 19.332 |
| Fe | 10,103.139 | 10,098.145 | 4.994 |
| Mn | 547.605 | 525.452 | 22.153 |
| Ni | 38.328 | 35.302 | 3.026 |
| Pb | 2305.09 | 2304.18 | 0.91 |
| Sb | 45.329 | 43.955 | 1.374 |
| Sn | 26.695 | 25.286 | 1.409 |
| Ti | 928.567 | 906.227 | 22.34 |
| V | 34.064 | 33.372 | 0.692 |
| Zn | 2047.331 | 2028.555 | 18.776 |
| Zr | 32.309 | 31.704 | 0.605 |

and kriging results should not be considered independently; instead, a combinatory model should be produced in the prediction of the geochemical elements for stream sediments. Further research should investigate if other geo-environmental factors such as proximity to the sea, structural geology, geophysical conditions, and terrain slope angle can be incorporated into the characterization model.

## Compliance with ethical standards

**Declarations**   Authors declare no conflict of interest.

## References

Arndt NT, Fontboté L, Hedenquist JW, Kesler SE, Thompson JF, Wood DG (2017) Section 3. Mineral exploration: discovering and defining ore bodies. Geochem Perspect 6:52–85

Bai J, Ng S (2005) Tests for skewness, kurtosis, and normality for time series data. J Bus Econ Stat 23:49–60

Beedles WL, Simkowitz MA (1978) A note on skewness and data errors. J Financ 33:288–292

Brown M (1993) Exploration of gold in central wales. British Geological Survey, Keyworth, Nottingham

Cai Y, Li J, Li X, Li D, Zhang L (2018) Estimating soil resistance at unsampled locations based on limited CPT data. Bull Eng Geol Environ 1–12

Carranza EJM (2017) Natural resources research publications on geochemical anomaly and mineral potential mapping, and introduction to the special issue of papers in these fields. Nat Resour Res 26:379–410

Cheng Q (2007) Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geol Rev 32:314–324

Cheng Q (2012) Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. J Geochem Explor 122:55–70

Cohen D, Silva-Santisteban C, Rutherford N, Garnett D, Waldron H (1999) Comparison of vegetation and stream sediment geochemical patterns in northeastern New South Wales. J Geochem Explor 66:469–489

Correia MIT, Waitzberg DL (2003) The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. Clin Nutr 22:235–239

Cressie NA (1993) Statistics for spatial data: Wiley series in probability and mathematical statistics Find this article online

Dagdelen K, Vega AN (1997) Geostatistics applied to mine waste characterization at Leadville, Colorado, USA. Int J Surf Min Reclam Environ 11:175–188

de Mulder EF, Cheng Q, Agterberg F, Goncalves M (2016) New and game-changing developments in geochemical exploration. Episodes 39:70–71

Deutsch CV, Journel AG (1992) GSLIB: Geostatistical Software Library and User's Guide. Hauptbd. Oxford university press

Ding Z, Mei G, Cuomo S, Xu N, Tian H (2017) Performance evaluation of gpu-accelerated spatial interpolation using radial basis functions for building explicit surfaces. Int J Parallel Prog 1–29

Facchinelli A, Sacchi E, Mallen L (2001) Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. Environ Pollut 114:313–324

Fletcher W (1997) Stream sediment geochemistry in today's exploration world. In: Proceedings of exploration. pp 249–260

Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228:113–129

Hattis D, Ogneva-Himmelberger Y, Ratick S (2012) The spatial variability of heat-related mortality in Massachusetts. Appl Geogr 33:45–52

Hou D, O'Connor D, Nathanail P, Tian L, Ma Y (2017) Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: a critical review. Environ Pollut 231:1188–1200

Johnston K, Hoef J, Krivoruchko K, Lucas N (2012) ArcGIS® 9: using ArcGIS® geostatistical analyst. ESRI

Jones R (2002) Algorithms for using a DEM for mapping catchment areas of stream sediment samples. Comput Geosci 28:1051–1060

Kent J, Leitner M (2007) Efficacy of standard deviational ellipses in the application of criminal geographic profiling. J Investig Psychol Offender Profiling 4:147–165

Khalajmasoumi M, Sadeghi B, Carranza EJM, Sadeghi M (2017) Geochemical anomaly recognition of rare earth elements using multi-fractal modeling correlated with geological features, Central Iran. J Geochem Explor 181:318–332

Khamis N, Sin TC, Hock GC (2018) Segmentation of residential customer load profile in peninsular Malaysia using Jenks natural breaks. In: 2018 IEEE 7th International Conference on Power and Energy (PECon). IEEE, pp 128–131

Kleinschmidt I, Bagayoko M, Clarke GPY, Craig M, Le Sueur D (2000) A spatial statistical approach to malaria mapping. Int J Epidemiol 29:355–361. https://doi.org/10.1093/Ije/29.2.355

Komnitsas K, Guo X, Li D (2010) Mapping of soil nutrients in an abandoned Chinese coal mine and waste disposal site. Miner Eng 23: 627–635

Kravchenko A, Bullock DG (1999) A comparative study of interpolation methods for mapping soil properties. Agron J 91:393–400

Lefever DW (1926) Measuring geographic concentration by means of the standard deviational ellipse. Am J Sociol 32:88–94

Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: a review. Environ Model Softw 53:173–189

Liu X, Wu J, Xu J (2006) Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS. Environ Pollut 141:257–264

Liu Y, Cheng Q, Carranza EJM, Zhou K (2018) Assessment of geochemical anomaly uncertainty through geostatistical simulation and singularity analysis. Nat Resour Res 1–14

Lu GY, Wong DW (2008) An adaptive inverse-distance weighting spatial interpolation technique. Comput Geosci 34:1044–1055

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266

Merritt WS, Letcher RA, Jakeman AJ (2003) A review of erosion and sediment transport models. Environ Model Softw 18:761–799

Moral FJ (2010) Comparison of different geostatistical approaches to map climate variables: application to precipitation. Int J Climatol 30:620–631

Mueller T, Pusuluri N, Mathias K, Cornelius P, Barnhisel R, Shearer S (2004) Map quality for ordinary kriging and inverse distance weighted interpolation. Soil Sci Soc Am J 68:2042–2047

Nieto A, Toffait Y (2007) Tonnage-grade test of nonlinear estimators: indicator kriging, disjunctive kriging and uniform conditioning. Trans Soc Min Metall Explor Inc 320:45

Nieto A, Bai Y, Brownson J (2014) Combined life cycle assessment and costing analysis optimization model using multiple criteria decision making in earth-resource systems. Nat Resour 5:351

Olea RA (2006) A six-step practical approach to semivariogram modeling. Stoch Env Res Risk A 20:307–318

Ottesen R, Bogen J, Bølviken B, Volden T (1989) Overbank sediment: a representative sample medium for regional geochemical mapping. J Geochem Explor 32:257–277

Panahi A, Cheng Q, Bonham-Carter GF (2004) Modelling lake sediment geochemical distribution using principal component, indicator kriging and multifractal power-spectrum analysis: a case study from Gowganda, Ontario. Geochem: Explor, Environ, Anal 4:59–70

Piel FB et al (2013) Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. Lancet 381:142–151

Płaczkowska E, Górnik M, Mocior E, Peek B, Potoniec P, Rzonca B, Siwek J (2015) Spatial distribution of channel heads in the Polish Flysch Carpathians. Catena 127:240–249

Reed GF, Lynn F, Meade BD (2002) Use of coefficient of variation in assessing variability of quantitative assays. Clin Diagn Lab Immunol 9:1235–1239

Robinson T, Metternicht G (2006) Testing the performance of spatial interpolation techniques for mapping soil properties. Comput Electron Agric 50:97–108

Salminen R, Gregorauskien V (2000) Considerations regarding the definition of a geochemical baseline of elements in the surficial materials in areas differing in basic geology. Appl Geochem 15:647–653

Seidl DE, Paulus G, Jankowski P, Regenfelder M (2015) Spatial obfuscation methods for privacy protection of household-level data. Appl Geogr 63:253–263

Shahbeik S, Afzal P, Moarefvand P, Qumarsy M (2014) Comparison between ordinary kriging (OK) and inverse distance weighted (IDW) based on estimation error. Case study: Dardevey iron ore deposit, NE Iran. Arab J Geosci 7:3693–3704

Ssempiira J, Nambuusi B, Kissa J, Agaba B, Makumbi F, Kasasa S, Vounatsou P (2017) Geostatistical modelling of malaria indicator survey data to assess the effects of interventions on the geographical distribution of malaria prevalence in children less than 5 years in Uganda. PLoS One 12:e0174948

Triantafilis J, Odeh I, McBratney A (2001) Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Sci Soc Am J 65:869–878

Varouchakis E, Hristopulos D (2013) Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. Environ Monit Assess 185:1–19

Voltz M, Webster R (1990) A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. J Soil Sci 41:473–490

Wang B, Shi WZ, Miao ZL (2015) Confidence Analysis of Standard Deviational Ellipse and Its Extension into Higher Dimensional Euclidean Space. PLoS One 10:e0118537. https://doi.org/10.1371/journal.pone.0118537

Wei Z, Xia W, Canjiong L, Renzhong S, Yuzhuo W (2016) Determination of flow accumulation threshold based on multiple regression model in raster river networks extraction. Tran Chin Soc Agric Mach 10:018

Wong DW (2016) Interpolation: inverse-distance weighting international encyclopedia of geography: people, the earth, environment and technology: people, the earth, environment and technology 1–7

Wu C, Wu J, Luo Y, Zhang H, Teng Y, DeGloria SD (2011) Spatial interpolation of severely skewed data with several peak values by the approach integrating kriging and triangular irregular network interpolation. Environ Earth Sci 63:1093–1103

Yang C-S, Kao S-P, Lee F-B, Hung P-S (2004) Twelve different interpolation methods: A case study of Surfer 8.0. In: Proceedings of the XXth ISPRS Congress. pp 778–785

Zuluaga MC, Norini G, Lima A, Albanese S, David CP, De Vivo B (2017) Stream sediment geochemical mapping of the Mount Pinatubo-Dizon Mine area, the Philippines: implications for mineral exploration and environmental risk. J Geochem Explor 175:18–35

Zuo R, Carranza EJM, Wang J (2016) Spatial analysis and visualization of exploration geochemical data. Earth Sci Rev 158:9–18