

A Multivariate Geostatistical Methodology to Delineate Areas of Potential Interest for Future Sedimentary Gold Exploration

Pierre Goovaerts¹ · M. T. D. Albuquerque² ·
I. M. H. R. Antunes²

Received: 27 September 2015 / Accepted: 31 December 2015 / Published online: 1 February 2016
© International Association for Mathematical Geosciences 2016

Abstract This paper describes a multivariate geostatistical methodology to delineate areas of potential interest for future sedimentary gold exploration, with an application to an abandoned sedimentary gold mining region in Portugal. The main challenge was the existence of only a dozen gold measurements confined to the grounds of the old gold mines, which precluded the application of traditional interpolation techniques, such as cokriging. The analysis could, however, capitalize on 376 stream sediment samples that were analysed for 22 elements. Gold (Au) was first predicted at all 376 locations using linear regression ($R^2 = 0.798$) and four metals (Fe, As, Sn and W), which are known to be mostly associated with the local gold's paragenesis. One hundred realizations of the spatial distribution of gold content were generated using sequential indicator simulation and a soft indicator coding of regression estimates, to supplement the hard indicator coding of gold measurements. Each simulated map then underwent a local cluster analysis to identify significant aggregates of low or high values. The 100 classified maps were processed to derive the most likely classification of each simulated node and the associated probability of occurrence. Examining the distribution of the hot-spots and cold-spots reveals a clear enrichment in Au along the Erges River downstream from the old sedimentary mineralization.

Keywords Sequential indicator simulation · Soft indicators · Cluster analysis · Linear regression · Accuracy plots

✉ Pierre Goovaerts
goovaerts@biomedware.com

¹ BioMedware Inc., PO Box 1577, Ann Arbor, MI 48106, USA

² CIGAR, Geo-Environmental and Resources Research Center, FEUP, Oporto and Polytechnic Institute of Castelo Branco, 6001-909 Castelo Branco, Portugal

1 Introduction

In Portugal, ore exploitation and processing has been an important economic activity, with open pit and underground structures, particularly until the early 1970s. The recent spike in metal prices and technologic developments have made extraction and processing more effective. In turn, this has led to renewed interest in previously abandoned gold mining areas, with a few experimental explorations undertaken in Alentejo, southern Portugal. A spatial approach is a suitable way to assess the potential for gold in old mining areas, as demonstrated in other countries (Darwish and Poellmann 2010).

Geochemistry is one of the most significant instruments for exploring undiscovered mineral resources (e.g., Cameron et al. 2004; Carranza et al. 2009; Carranza 2011; Hronsky 2004; Hronsky and Groves 2008; McCuaig et al. 2010; Wang et al. 2008). Geochemical cartography and the identification of associated anomalies have been a goal in mining prospecting techniques since the late 1920s with multiple applications to the estimation of gold mineralized deposits (Bin 1995; Goovaerts et al. 2014; Madani 2011; Viladevall et al. 1999). Recent development of analytical methods and computational resources facilitates the implementation of geochemical mapping and its use in natural resources management (Antunes and Albuquerque 2013). In particular, multivariate data analysis has been widely applied to characterize the statistical patterns of geochemical data (Carranza 2010; El-Makky 2011; Sadeghi et al. 2013, 2014; Viladevall et al. 1999; Zuo 2011). These methods commonly aim to reduce the dimensionality of the problem through the creation of a few relevant factors that explain a large proportion of variance in a multivariate data set (Davis 2002; Reimann et al. 2008).

Stream sediment surveys remain the common geochemical approach used for regional gold exploration where slope designs distinct drainage systems (Darwish and Poellmann 2010; Fletcher 1997; Hale and Plant 1994; Goovaerts et al. 2014). The exploration of sedimentary gold (alluvial) has been conducted in distinct mineralized areas and linked up with different genetic deposits (Chapman et al. 2000; Chapman and Mortensen 2006; McInnes et al. 2008; Mortensen et al. 2004; Outridge et al. 1998; Townley et al. 2003; Viladevall et al. 1999). The alluvial gold concentrates occur downstream from the ore deposits (McInnes et al. 2008). Lithological and structural criteria are considered to be the most important exploration criteria for gold deposits (Madani 2011).

The aim of this manuscript was the development and application of a spatial statistical approach for sedimentary gold exploration, with a focus on the visualization and delineation of potential zones of low and high values for future prospectings, instead of the accurate estimation of gold content. The methodology is illustrated using trace elements—Ag, As, B, Ba, Be, Cd, Co, Cr, Cu, Fe, Nb, Ni, Mn, Mo, Pb, Sb, Sn, V, Y, U, W and Zn—measured in 376 stream sediments samples that were collected in the old sedimentary gold abandoned mining region of Monfortinho (Central Portugal).

Section 2 describes the study area and the data available for modeling. The methodology is introduced in Sect. 3 and the results of its application to an old gold mine are discussed in Sect. 4. Conclusions are summarized in Sect. 5.

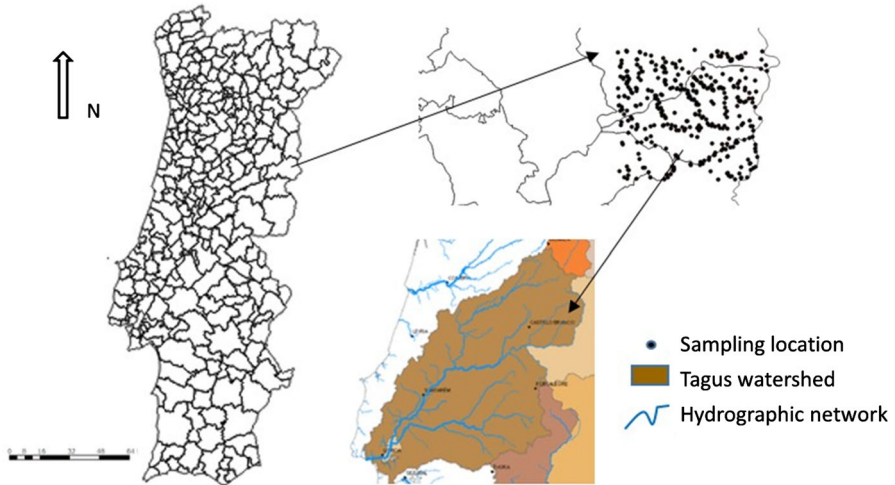


Fig. 1 Location of the study area in Portugal and sampling data configuration in the Tagus watershed

2 Study Area

The study area is located in Monfortinho region, about 70 km East of Castelo Branco (central Portugal), and is part of the Central Iberian Zone, in the Portuguese–Spanish border (Fig. 1). It is occupied mainly by the Cambrian schist-greywacke complex associated with Ordovician quartzites and covered by Tertiary sedimentary materials (Oliveira et al. 1992). The transboundary Portuguese–Spain border is delimited by the Erges River, one of the last wild rivers in Portugal, with a rare natural value due to its geodiversity. Agriculture is the main local activity and the thermal SPA water from Fonte Santa contributes to the economy and tourism of the region. This region is characterized by a dry climate with most streams drying up in the summer (Antunes et al. 2002).

The area of sampling site is approximately 140 km² and surrounds the mining of sedimentary gold in the Erges River. Around mine tailing sites, the mineralogical content of the material exploited consists of inert materials from the gangue constituent's mineralization or mineral constituents of rocks (Maroto et al. 1997). Geochemical anomalies found in the vicinity of tailings and mineralized areas indicate the action of dominant wind and transport of fine dust from the superficial layers of the heap (Santos Oliveira et al. 1998). The stream sediments, resulting from the alteration of rocks by various physical and chemical processes, are mobilized, transported and deposited along the water lines.

The geochemical composition of stream sediments and their spatial distribution in the study area were characterized using a total of 376 representative samples, collected between 1980 and 1988, in a narrow region ranging from 50 m upstream to 100 m downstream from the streams' confluences (Instituto Geológico e Mineiro 1988) (Fig. 1). Since almost all water lines correspond to open valleys, our point-support stream sediments samples correspond to incipient instead of evolved soils. All

the samples were collected on schist and their preparation included reduction, drying and grinding. Total concentration of Ag, As, Au, B, Ba, Be, Cd, Co, Cr, Cu, Fe, Nb, Ni, Mn, Mo, Pb, Sb, Sn, V, Y, U, W and Zn were analyzed by ICP-AES, with a precision of 20 % for As and 10 % for the other elements (Instituto Geológico e Mineiro 1988). Gold was measured in 12 samples collected inside the old mine area. Tin and W were analysed by X-ray fluorescence spectrometry and plasma emission with a precision of 10 % (0.05 ppb) (Antunes et al. 2002; Instituto Geológico e Mineiro 1988).

3 Methodology

The flowchart in Fig. 2 describes the different steps of the geospatial analysis which includes linear regression, indicator kriging (IK), sequential indicator simulation and local cluster analysis (LCA).

3.1 Hard and Soft Indicator Coding

Because gold content was measured only for a small subset ($n_1 = 12$) of stream sediment samples, the first step was to capitalize on the relationships between gold content and four metals (Fe, As, Sn and W), which are known to be mostly associated with the local gold’s paragenesis, to predict gold content at the remaining sampled locations ($n_2 = 364$). Although this relationship was established from in-situ gold content it is relevant to the 364 stream sediment samples since the geochemistry of Au and the other four elements are stable during mobility and weathering (Antunes et al. 2002; Harraz et al. 2012). This prediction was here based on linear regression, resulting in an estimated value $m_{LR}(\mathbf{u}_\alpha)$ and associated standard error $s_{LR}(\mathbf{u}_\alpha)$ at all n_2 locations with geographical coordinates \mathbf{u}_α .

To account for the uncertainty attached to n_2 gold content estimates (soft data) relative to n_1 measurements (hard data), each of the 12 hard data $z(\mathbf{u}_\alpha)$ and 364 soft data $m_{LR}(\mathbf{u}_\alpha)$ were transformed into a set of $K = 19$ indicators as follows

$$i(\mathbf{u}_\alpha; z_k) = \begin{cases} 1 & \text{if } z(\mathbf{u}_\alpha) \leq z_k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$i(\mathbf{u}_\alpha; z_k) = G\left(\frac{z_k - m_{LR}(\mathbf{u}_\alpha)}{s_{LR}(\mathbf{u}_\alpha)}\right), \tag{2}$$

where z_k is a gold content threshold identified with 5kth percentile of the distribution of hard and soft data, and $G(\cdot)$ is the standard normal cumulative distribution function. The spatial connectivity of these indicators was then quantified and modeled using the indicator semivariogram defined as

$$\gamma(\mathbf{h}; z_k) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [i(\mathbf{u}_\alpha; z_k) - i(\mathbf{u}_\alpha + \mathbf{h}; z_k)]^2, \tag{3}$$

where the indicator value at location \mathbf{u}_α is paired with another indicator value a lag distance \mathbf{h} away.

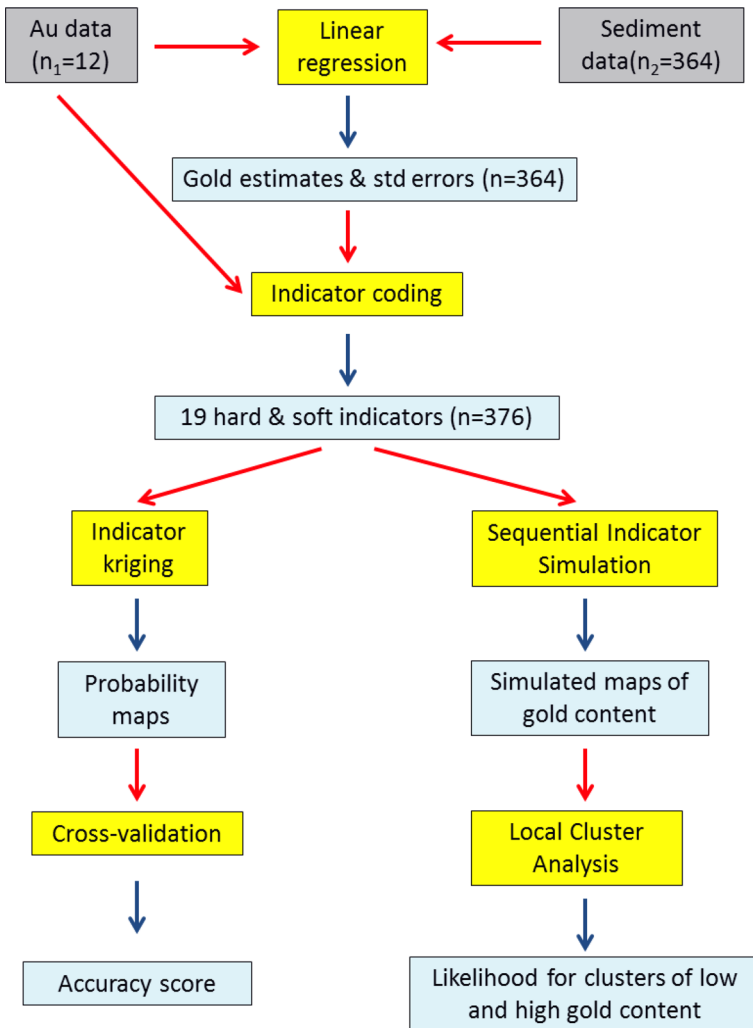


Fig. 2 Flowchart describing the different steps of the analysis conducted on 12 gold data and 376 stream sediment metal data to delineate areas of low and high gold content

3.2 Indicator Kriging and Cross-Validation

The delineation of zones of low and high values required the interpolation of gold content to the nodes of a regular grid. This step was accomplished using soft indicator kriging (Goovaerts 1994) whereby the probability of being no greater than a threshold z_k at any node \mathbf{u}_0 was estimated as the following linear combination of indicators

$$F_{IK}(\mathbf{u}_0; z_k | (n)) = \sum_{\alpha=1}^{n(\mathbf{u}_0)} \lambda_{\alpha k} i(\mathbf{u}_\alpha; z_k) \quad k = 1, \dots, K, \quad (4)$$

where $\lambda_{\alpha k}$ are kriging weights that are solutions of the following system of linear equations

$$\begin{cases} \sum_{\alpha=1}^{n(\mathbf{u}_0)} \lambda_{\alpha k} C_I(\mathbf{u}_\alpha - \mathbf{u}_\beta; z_k) + \mu(\mathbf{u}_\beta; z_k) = C_I(\mathbf{u}_\beta - \mathbf{u}_0; z_k) \\ \sum_{\alpha=1}^{n(\mathbf{u}_0)} \lambda_{\alpha k} = 1 \end{cases} \quad \beta = 1, \dots, n(\mathbf{u}_0) . \tag{5}$$

The indicator covariance function $C_I(\mathbf{h}; z_k)$ was derived by subtracting the model fitted to the experimental indicator semivariogram (Eq. 3) from the sill. Indicator kriging was conducted using the program AUTO-IK (Goovaerts 2009) which computes at each node \mathbf{u}_0 the mean and variance of the local distributions of probability (ccdf), denoted $\hat{z}_E(\mathbf{u}_0)$ (E-type estimate) and $s_E^2(\mathbf{u}_0)$, after correction of K probability estimates (Eq. 4) for order relation deviations and interpolation/extrapolation to complete the discrete ccdf.

The quality of the model of uncertainty provided by indicator kriging was assessed using a leave-one-out cross-validation approach whereby IK results at sampled locations \mathbf{u}_α were compared to observations (soft or hard data) that were removed one at a time. From the ccdf $F_{IK}(\mathbf{u}_\alpha; z|(n))$ one can compute a series of symmetric median-centred p -probability intervals (PI) bounded by the $(1 - p)/2$ and $(1 + p)/2$ quantiles of that distribution: $q(\mathbf{u}_\alpha; (1 - p)/2)$ and $q(\mathbf{u}_\alpha; (1 + p)/2)$. For example, the 0.5-PI is bounded by the lower and upper quartiles of the ccdf. According to this model of local uncertainty, there is then a 0.5-probability that the actual attribute value (i.e., gold content) falls into the 0.5-PI or, equivalently, that over the study area 50 % of the 0.5-PI includes the true z values. The fraction of true values falling into the symmetric p -PI was here computed as

$$p^* = \frac{1}{n_1} \sum_{\alpha=1}^{n_1} j(\mathbf{u}_\alpha; p) + \frac{1}{n_2} \left[\sum_{\alpha=1}^{n_2} G \left(\frac{q(\mathbf{u}_\alpha; (1 + p)/2) - m_{LR}(\mathbf{u}_\alpha)}{s_{LR}(\mathbf{u}_\alpha)} \right) - G \left(\frac{q(\mathbf{u}_\alpha; (1 - p)/2) - m_{LR}(\mathbf{u}_\alpha)}{s_{LR}(\mathbf{u}_\alpha)} \right) \right], \tag{6}$$

where $j(\mathbf{u}_\alpha; p) = 1$ if the hard data $z(\mathbf{u}_\alpha)$ falls within the p -PI, and zero otherwise. For the n_2 soft data, the overlap between the p -PI and the Gaussian distribution centred on the regression estimate $m_{LR}(\mathbf{u}_\alpha)$ is computed using an expression similar to Eq. (2). Following Deutsch (1997), the agreement between observed, p_k^* , and expected fractions, p_k , is quantified using the following “goodness” statistic

$$G = 1 - \frac{1}{K'} \sum_{k=1}^{K'} w_k |p_k^* - p_k| \quad \text{with } 0 \leq G \leq 1, \tag{7}$$

where $w_k = 1$ if $p_k^* > p_k$, and 2 otherwise. K' represents the discretization level of the computation. Twice more importance is given to deviations when $p_k^* < p_k$ (inaccurate

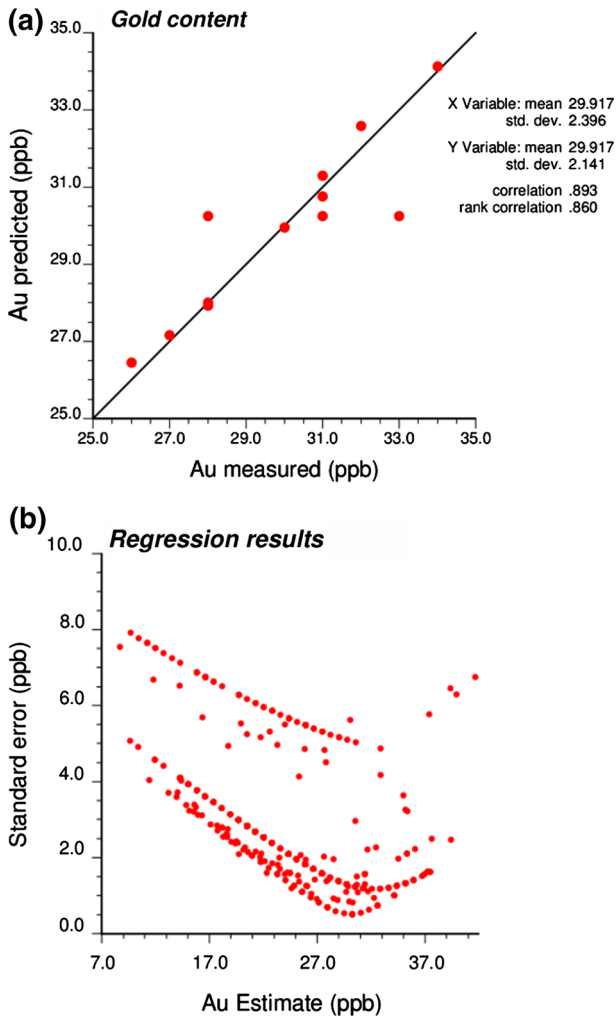


Fig. 3 **a** Scatterplot of measured gold content ($n_1 = 12$) versus estimates computed by linear regression from four metals (Fe, As, Sn and W). **b** Results of the application of this regression model to 364 stream sediment samples. *Bottom scatterplot* illustrates the larger uncertainty (standard error) associated to values extrapolated beyond the range (26 to 34 ppb) of the 12 gold measurements

case). The weights penalize less the accurate case, which is the case where the fraction of true values falling into the p -probability interval is larger than expected. The goodness statistic is completed by the so-called “accuracy plot” that allows one to visualize departures between observed and expected fractions as a function of the probability p .

3.3 Local Cluster Analysis

The delineation of zones of low and high contents in gold was conducted through the application of local cluster analysis (Anselin 1995; Fu et al. 2014; Goovaerts

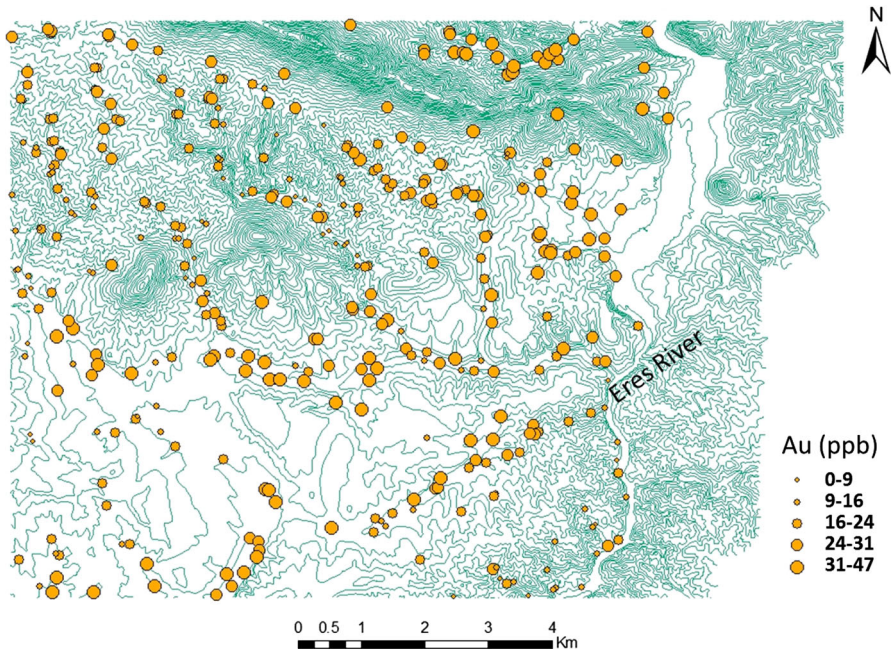


Fig. 4 Location map of stream sediment samples overlaid on a topographic map of the study area. The size of yellow dots is proportional to the gold concentration estimated by linear regression

et al. (2005a, b); Goovaerts (2010)). The basic idea is to compute at each grid node a local indicator of spatial autocorrelation (LISA) and test whether this statistic is significantly positive, indicating the existence of an aggregate of grid nodes with similar gold content, either low or high. Similarity between the gold concentration E-type estimate at node \mathbf{u}_0 (kernel value) and values estimated at $J(\mathbf{u}_0)$ adjacent nodes \mathbf{u}_j (e.g., $J(\mathbf{u}_0) = 8$ nodes adjacent to \mathbf{u}_0) was here quantified by the local Moran's I statistic defined as

$$I(\mathbf{u}_0) = \frac{\hat{z}_E(\mathbf{u}_0) - m}{s} \times \frac{1}{J(\mathbf{u}_0)} \sum_{j=1}^{J(\mathbf{u}_0)} \frac{\hat{z}_E(\mathbf{u}_j) - m}{s}, \quad (8)$$

where m and s are the mean and standard deviation of the set of N grid estimates. This local statistic is simply the product of the kernel value and the average of neighboring values; it can detect both positive and negative autocorrelations. It exceeds zero if the kernel and neighborhood averaged gold content estimates jointly exceed the global mean m (High–High, HH cluster) or are jointly below m (Low–Low, LL cluster). LISA values are negative if the kernel and neighborhood mean values are on opposite sides of the global mean m , which indicates the presence of spatial outliers or anomalies: High–Low (HL outlier) or Low–High (LH outlier).

To test whether any test statistic, $I(\mathbf{u}_0)$, is significantly greater or smaller than 0 (i.e., presence of spatial autocorrelation), one needs to know its probability distribution under the null hypothesis of spatial independence. The common way to generate such

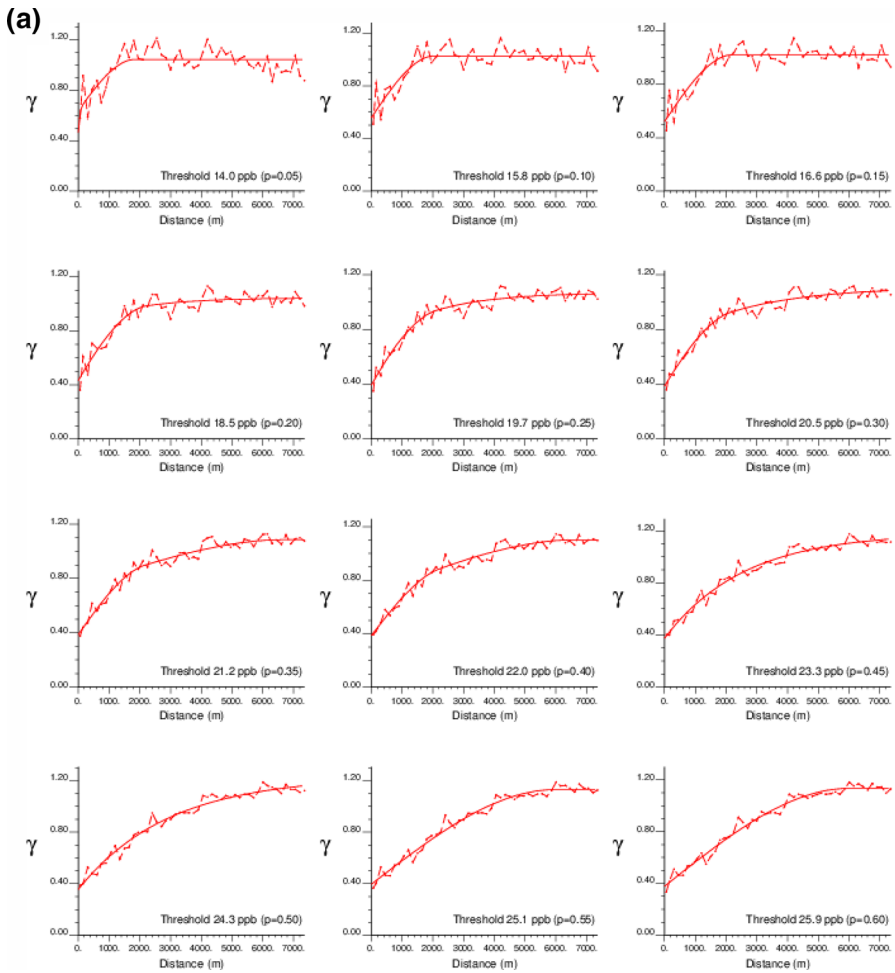


Fig. 5 Experimental indicator semivariograms standardized to unit sill with the isotropic model fitted automatically by AUTO-IK for each of the 19 gold content thresholds

reference distribution is to shuffle the set of estimated values randomly and then to use the shuffled values to compute the neighborhood average in Eq. (8) while the kernel gold content remains the same. In other words, the LISA statistic is computed for randomly distributed gold contents in adjacent locations. This operation is repeated K times ($K = 999$ in this article) to compute the P value of the test. Because the statistical test is repeated for each grid node, there is an increased likelihood of false positives (i.e., risk of rejecting the null hypothesis when it is true) and the test needs to be corrected for multiple testing. This correction was here accomplished using the false discovery rate (FDR) approach, which aims to control the expected proportion of true null hypotheses that will be rejected (Castro and Singer 2006); that is the objective is to limit the risk of false positives.

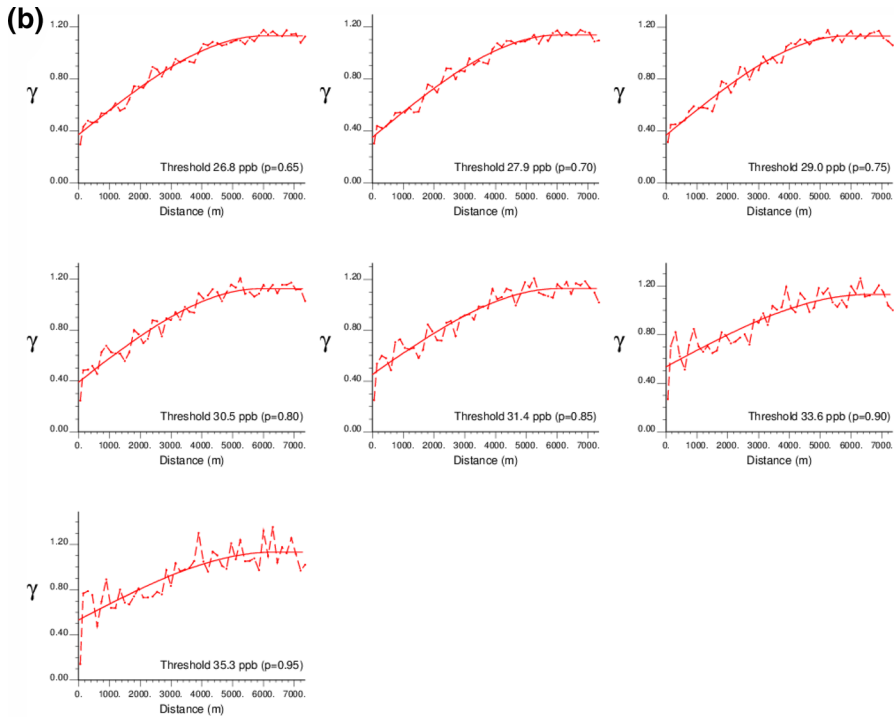
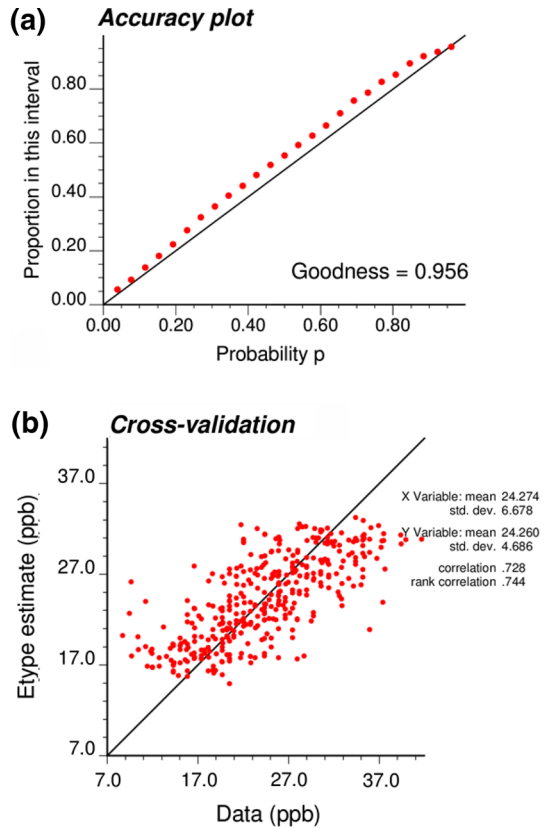


Fig. 5 continued

3.4 Propagation of Uncertainty Using Sequential Indicator Simulation

The application of local cluster analysis to E-type estimates has two main drawbacks: (i) the detection of artificial clusters resulting from the autocorrelation imparted to estimates $\hat{z}_E(\mathbf{u}_0)$ by the smoothing effect of kriging, and (ii) the failure to account for the uncertainty attached to kriging estimates (e.g., variance $s_E^2(\mathbf{u}_0)$ of the ccdf). Goovaerts (2006) proposed a simulation-based approach to account for uncertainty in local cluster analysis and avoid the smoothing effect of kriging. First, the uncertainty attached to the spatial distribution of gold content is modeled through the generation of a set of equally-probable simulated maps, $\{z^{(l)}(\mathbf{u}_j), j = 1, \dots, N; l = 1, \dots, L\}$, each consistent with the information available, such as histogram or a spatial correlation function. This step was here accomplished using sequential indicator simulation based on soft indicator coding (Sect. 3.1). Then, the uncertainty is propagated through the computation of the LISA statistic by replacing in Eq. (8) the E-type estimates $\hat{z}_E(\mathbf{u}_0)$ and $\hat{z}_E(\mathbf{u}_j)$ by the corresponding simulated values, leading to a set of L simulated LISA values $\{I^{(l)}(\mathbf{u}_0), l = 1, \dots, L\}$. In other words, the correlation of each node with adjacent nodes will be tested L times, enabling the computation of the probability for that node to belong to a cluster of small gold content (low value surrounded by low values) or a cluster of large gold content (high value surrounded by high values). The L classified maps are then processed to derive the most likely classification

Fig. 6 Results of a cross-validation of soft indicator kriging: **a** accuracy plot, and **b** scatterplot of E-type estimates (ccdf mean) versus gold data used for hard and soft indicator coding and displayed in Fig. 4



of each node and the associated likelihood (i.e., frequency of occurrence of that class over L simulations).

4 Results and Discussion

A linear regression analysis of gold content versus four metals (Fe, As, Sn and W) explained close to 80 % of the total variance ($R^2 = 0.798$). Although the large R^2 is due to some extent to the small number of data available ($n = 12$), the scatterplot in Fig. 3a indicates a good agreement between the predicted and measured gold concentrations. The application of this regression model to the remaining 364 stream sediment samples led to gold content estimates that are plotted versus their standard error in Fig. 3b. This scatterplot clearly illustrates the larger uncertainty (standard error) associated with values extrapolated beyond the range (26 to 34 ppb) of the twelve gold measurements. Such an extrapolation makes sense in the present exploratory setting where few gold samples are available and lower values are expected away from the location of the old gold mine. The larger uncertainty attached to the prediction of these lower values is incorporated into the analysis through the soft indicator coding and the subsequent

Ccdf standard deviation

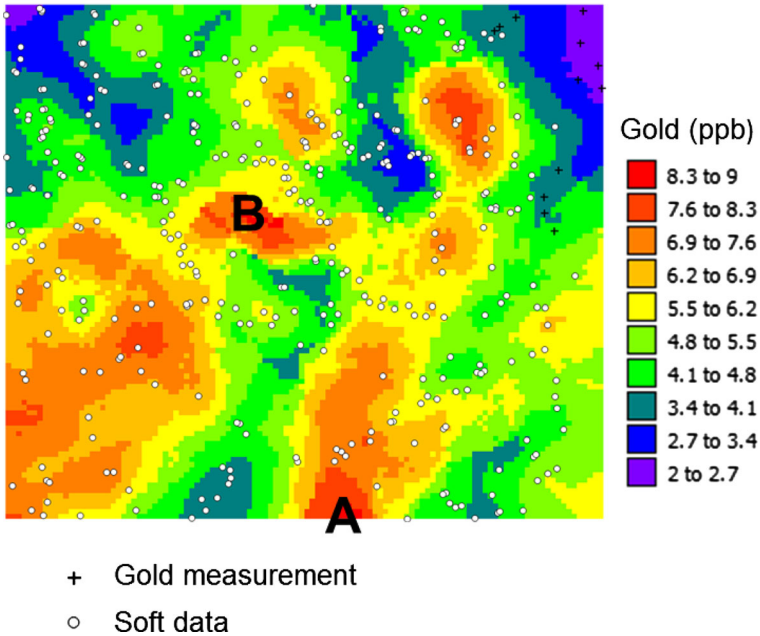


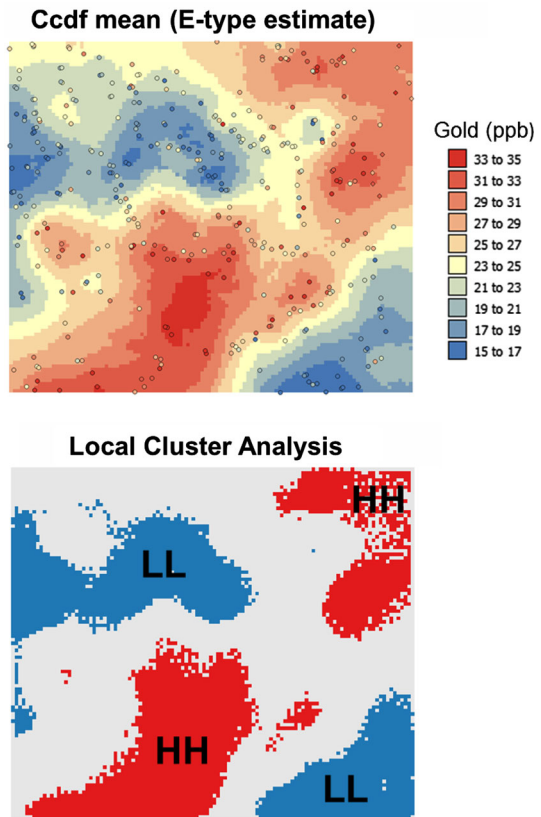
Fig. 7 Map of the standard deviation of the conditional cumulative distribution functions (ccdf) derived using soft indicator kriging. *Black crosses* denote the location of the 12 hard data (gold content) measured inside the abandoned gold mines whereas *white dots* are the soft data

simulation procedure. The gold content estimated by regression is mapped in Fig. 4. Zones of low and high gold contents could, however, not be readily delineated from this location map because of the discrete nature of the sampling and the fact that the uncertainty attached to the regression estimates was ignored.

Each of the 12 gold data and 364 regression estimates were coded into a set of 19 indicators according to Eqs. (1) and (2). This high level of discretization was justified by the fact that each soft data take the form of a Gaussian probability distribution centred on the regression estimate. The thresholds were identified from the histogram of 376 values in order to split the sample distribution into 20 classes of equal frequency. The indicator semivariograms in Fig. 5 indicate a good spatial connectivity of the indicators regardless the threshold. In particular the average relative nugget effect is 38.2 %, well within the range of 25 to 75 % commonly used to characterize a moderate spatial dependence (Cambardella et al. 1994). Interestingly the range of autocorrelation increases with the threshold: it is around 2 km for the smallest threshold's concentrations (<17 ppb) and exceeds 6 km for the largest thresholds (>23 ppb). In other words, the highest gold values are better connected in space than the lowest gold values, which is unlike most environmental datasets where high concentrations are often isolated hotspots (e.g., Goovaerts et al. 2005a, b).

Indicator semivariogram models were used with ordinary kriging to interpolate hard and soft indicator data to the nodes of a grid with 100 m spacing. A small per-

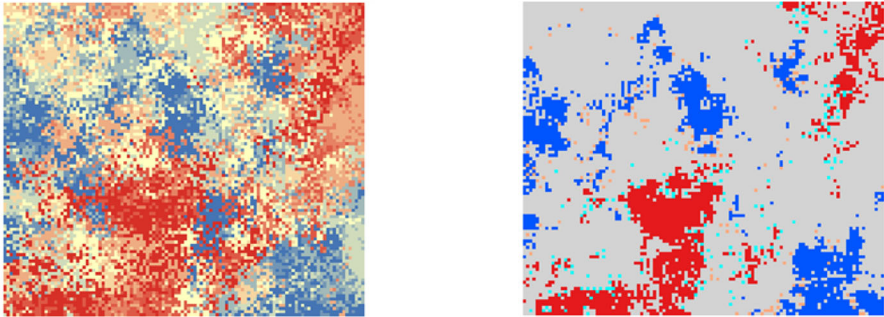
Fig. 8 Results of a local cluster analysis conducted on the map of gold content estimated by soft indicator kriging: significant clusters of low (LL) and high (HH) gold content detected using local Moran's *I*. *Dots* correspond to sampled locations



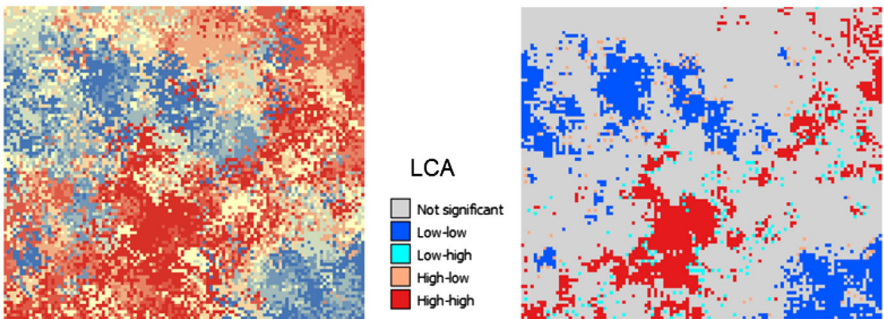
centage (1.9 %) of kriged probabilities had to be slightly corrected (average correction = 0.0015) in order to create valid cumulative probability distributions at each of the 11,286 grid nodes. The resolution of the discrete ccdf was then increased by performing a linear interpolation between tabulated bounds provided by the sample histogram (Deutsch and Journel 1998). The accuracy of the resulting model of uncertainty was first quantified using cross-validation. The accuracy plot indicates a very good agreement between observed and expected proportions of true values falling into probability intervals, leading to a goodness statistics close to 1 (Fig. 6a). The mean absolute prediction error was 3.6 ppb and the scatterplot in Fig. 6b illustrates the good correlation between E-type estimates and the original 376 data ($r = 0.73$). As for all least-squares interpolation methods (Goovaerts 1997), results display a conditional bias whereby the large concentrations are underestimated while the small concentrations are overestimated.

Figure 7 shows the map of the standard deviation of the local distributions of probability (ccdf) which can be interpreted as a measure of local uncertainty. As expected, the uncertainty is the lowest within the gold mines (upper right corner) where actual gold content was measured at 12 locations denoted by black crosses. At other locations the uncertainty combines the standard error of regression estimates where

Realization # 1



Realization # 2



Realization # 3

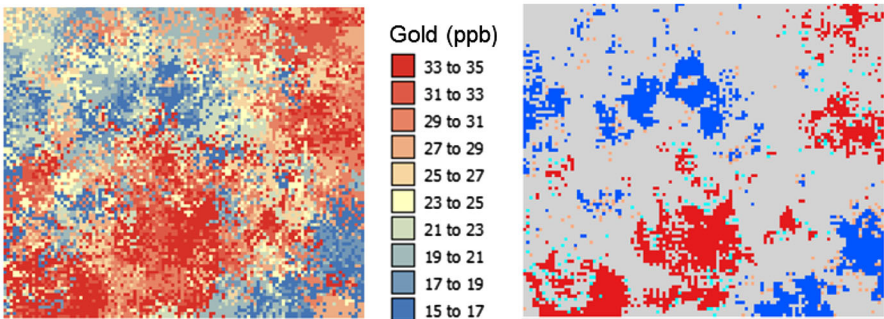
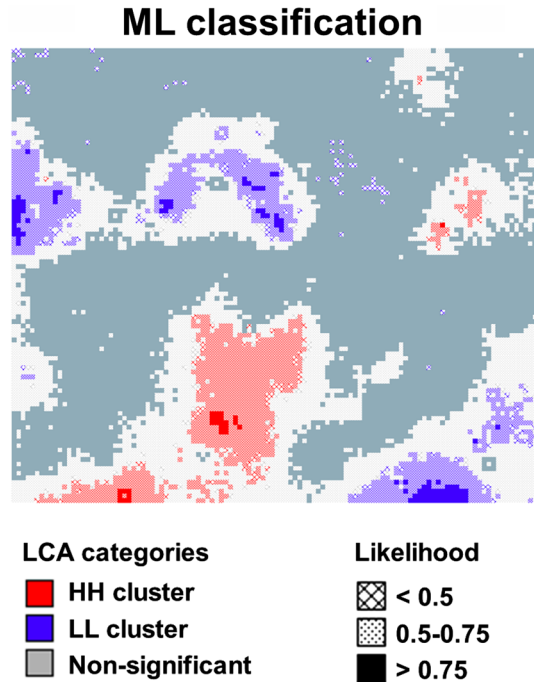


Fig. 9 Three simulated maps of gold content created using sequential indicator simulation (*left*) and the corresponding results of local cluster analysis (*right*)

gold was not recorded (white dots) with the uncertainty caused by spatial interpolation. The uncertainty is particularly large in sparsely sampled areas (zones A and B) and in zones of greater spatial variability which border between areas of low and high gold content displayed in the map of E-type estimates at the top of Fig. 8. A local cluster analysis was conducted using a significance level $\alpha = 0.05$ and the false

Fig. 10 Map of the most frequent cluster category (LL, HH, non-significant) computed from a local cluster analysis (LCA) of one hundred simulated maps of gold content. The intensity of the shading is proportional to the likelihood of occurrence of that category



discovery rate (FDR) approach for multiple testing correction. Two significant clusters of low gold content (LL) and two significant clusters of high gold content (HH) were identified within the study area (Fig. 8, bottom). Of particular interest are the higher gold content estimates found along the Erges River and downstream from the old abandoned sedimentary mineralization.

A better alternative to the smooth map of E-type estimates is an ensemble of simulated maps which reproduce the variability displayed by hard and soft data and model the uncertainty attached to their spatial distribution. One hundred maps were generated by sequential indicator simulation and each underwent a local cluster analysis similar to the one conducted for the single map of E-type estimates. Figure 9 shows the results for the first three simulated maps, which illustrates the uncertainty attached to the spatial distribution of gold content and how it impacts the definition of the LL and HH clusters. One obvious difference with kriging results (Fig. 8, bottom) is the much smaller size and spatial compactness of the clusters of low and high gold content; in particular, the cluster of high gold content immediately downstream from the gold mines vanished almost completely on some simulations. The remaining HH cluster is, however, bigger and more spatially compact than LL clusters, which agrees with the longer range of autocorrelation displayed by the indicator semivariograms corresponding to higher thresholds (Fig. 5). This difference in size between LL and HH clusters was less apparent on the classification map based on E-type estimates because the smoothing effect of kriging artificially inflated the size of these clusters.

The local Moran's *I* was also found significantly lower than zero at multiple locations, which indicates the presence of spatial outliers or anomalies. The set of 100 LCA results was summarized by assigning each grid node to the most frequent category (ML classification) and reporting that maximum frequency. Figure 10 confirms the conclusions drawn from the first three realizations regarding the severe shrinking of the cluster of high gold content located immediately downstream from the gold mines: only one node has a likelihood of belonging to that cluster above 0.75 and even for a 0.5 likelihood the cluster is pretty small with a size less than 5 % the size of the same cluster delineated on the basis of E-type estimates (Fig. 8, bottom). In comparison, the HH cluster in the Southern part of the study area represents 35 % of the counterpart kriging HH cluster and is now split into two parts. For both HH and LL clusters, the centres of the cluster are typically associated with the highest likelihood (i.e., more intense shading). Note that none of the nodes in the ML classification is flagged as a significant outlier.

5 Conclusions

This paper presented a multivariate geostatistical methodology to delineate areas of potential interest for future sedimentary gold exploration. The challenge was the existence of only a dozen of gold measurements confined to the grounds of the old gold mines, which precluded the application of traditional interpolation techniques, such as cokriging. The spatial characterization of the study area could, however, rely on the availability of a large set of stream sediment samples and the relationship between concentrations of several metals and gold content that was modeled using linear regression. This information translated into a set of prior distributions of probability that were discretized into indicator vectors and their semivariograms revealed the stronger spatial connectivity of larger gold concentration estimates (>23 ppb) relative to smaller concentrations (<17 ppb).

Soft indicator kriging allowed the derivation of the distributions of probability for gold content at the nodes of a regular grid, accounting for the uncertainty caused by the initial linear regression, in particular when extrapolating beyond the range of the twelve gold data, and the uncertainty resulting from the spatial interpolation. The use of the public-domain program AUTO-IK (Goovaerts 2009) greatly facilitated the application of indicator kriging with 19 thresholds as all the computation, including the modeling of 19 indicator semivariograms, was done automatically. Cross-validation demonstrated the accuracy of these models of uncertainty. The delineation of aggregates of low and high gold content was accomplished using local cluster analysis (LCA), which is commonly used to detect cancer clusters (Goovaerts and Jacquez 2005), yet has been seldom used in earth sciences (Zhang et al. 2008). Unlike a simple visual interpretation of a map of kriging estimates, LCA is based on statistical testing of the strength of spatial autocorrelation and allows the delineation of zones of significant clustering. One main difference between the applications of LCA in earth sciences relative to epidemiology is the sheer size of the dataset (e.g., large interpolation grid versus a few hundred administrative units), which increased the risk of false detection

through multiple testing. This issue was here addressed using the false discovery rate (FDR) approach.

Because it is a least-square interpolator, kriging tends to create smooth maps which display greater spatial continuity than the original data and a conditional bias. This was a shortcoming for the application of local cluster analysis since it led to artificially large clusters. In addition, the model of local uncertainty provided by the ccdf, which captures the uncertainty attached to applying regression outside the range of the data, was useless to characterize the uncertainty of a multi-point statistic like the local Moran's I. Both drawbacks were overcome by implementing the approach within a stochastic simulation framework. Sequential indicator simulation generated maps that reproduce the variability displayed by gold data, resulting in the identification of clusters of smaller size and the detection of spatial anomalies or outliers corresponding to significant negative spatial auto correlation. On the other end, the ensemble of 100 realizations provided a model of spatial uncertainty that could be propagated through the local cluster analysis to compute the likelihood of the clusters of high gold concentrations. In the current application, we found a clear Au enrichment along the Erges River downstream from the old abandoned sedimentary mineralization. The likelihood of this cluster of higher gold content would justify the future sampling of this area to validate the findings of the present study. The local likelihood could also help prioritizing which locations should be explored first, information that would be unavailable from a simple visual interpretation of maps of E-type estimates.

Acknowledgments We are grateful to Instituto Geológico e Mineiro (Portugal) for the data on stream sediments and associated informations about the study area. The research by P. Goovaerts was funded by Grant R21 ES021570 from the National Cancer Institute. The views stated in this publication are those of the author and do not necessarily represent the official views of the NCI.

References

- Anselin L (1995) Local indicators of spatial association—LISA. *Geogr Anal* 27:93–115
- Antunes IMHR, Albuquerque MTD (2013) Using indicator kriging for the evaluation of arsenic potential contamination in an abandoned mining area (Portugal). *Sci Total Environ* 442:545–552
- Antunes IMHR, Neiva AMR, Silva MMVG (2002) The mineralized veins and the impact of old mine workings on the environment at Segura, central Portugal. *Chem Geol* 190:417–31
- Bin H (1995) Assessment of geochemical anomalies for gold exploration using lead isotopes in west Hainan, China. *J Geochem Explor* 55:33–48
- Cambardella CA, Moorman AT, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field-scale variability of soil properties in central Iowa soils. *Soil Sci Soc Am J* 58:1501–1511
- Cameron EM, Hamilton SM, Leybourne MI, Hall GEM, McClenaghan MB (2004) Finding deeply buried deposits using geochemistry. *Geochem Explor Environ Anal* 4:7–32
- Carranza EJM (2010) Mapping of anomalies in continuous and discrete fields of stream sediment geochemical landscapes. *Geochem Explor Environ Anal* 10:171–187
- Carranza EJM (2011) Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. *J Geochem Explor* 110:167–185
- Carranza EJM, Owusu EA, Hale M (2009) Mapping of prospectivity and estimation of number of undiscovered prospects for lode gold, southwestern Ashanti Belt, Ghana. *Miner Depos* 44:915–938
- Castro MC, Singer BH (2006) Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. *Geogr Anal* 38:180–208
- Chapman RJ, Mortensen JK (2006) Application of microchemical characterization of placer gold grains to exploration for epithermal gold mineralization in regions of poor exposure. *J Geochem Explor* 91:1–26

- Chapman RJ, Leake RC, Moles RN, Earls G, Copper C, Harrington K, Berzins R (2000) The application of microchemical analysis of alluvial gold grains to the understanding of complex local and regional gold mineralization: a case study in the Irish and Scottish Caledonides. *Econ Geol Bull Soc Econ Geol* 95:1753–1773
- Darwish MAG, Poellmann H (2010) Geochemical exploration for gold in the Nile Valley Block (A) area, Wadi Allaqi, South Egypt. *Chem Erde* 70:353–362
- Davis JC (2002) *Statistics and data analysis in geology*. Wiley, New York 638 pp
- Deutsch CV (1997) Direct assessment of local accuracy and precision. In: Baafi EY, Schofield NA (eds) *Geostatistics wollongong '96*. Kluwer, Dordrecht, pp 115–125
- Deutsch CV, Journel AG (1998) *GSLIB: geostatistical software library and user's guide*, 2nd edn. Oxford University Press, New York 369 p
- El-Makky AM (2011) Statistical analyses of La, Ce, Nd, Y, Nb, Ti, P and Zr in bedrocks and their significance in geochemical exploration at the Um Garayat gold mine Area, Eastern Desert, Egypt. *Nat Resour Res* 20:157–176
- Fletcher WK (1997) Stream sediment geochemistry in today's exploration world. In: Gubins AG (ed) *Proceedings of exploration 97—4th decennial international conference on mineral exploration*, pp 249–260
- Fu WJ, Jiang PK, Zhou GM, Zhao KL (2014) Using Moran's and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of south-eastern China. *Biogeosciences* 11:2401–2409
- Goovaerts P (1994) Comparative performance of indicator algorithms for modeling conditional probability distribution functions. *Math Geol* 26(3):389–411
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, New York
- Goovaerts P (2006) Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson kriging and p-field simulation. *Int J Health Geogr* 5:7
- Goovaerts P (2009) AUTO-IK: a 2D indicator kriging program for the automated non-parametric modeling of local uncertainty in earth sciences. *Comput Geosci* 35:1255–1270
- Goovaerts P (2010) Geostatistical analysis of county-level lung cancer mortality rates in the Southeastern US. *Geogr Anal* 42:32–52
- Goovaerts P, Jacquez GM (2005) Detection of temporal changes in the spatial distribution of cancer rates using local Moran's I and geostatistically simulated spatial neutral models. *J Geogr Syst* 7:137–159
- Goovaerts P, AvRuskin G, Meliker J, Slotnick M, Jacquez GM, Nriagu J (2005) Geostatistical modeling of the spatial variability of arsenic in groundwater of Southeast Michigan. *Water Resour Res* 41(7):W07013
- Goovaerts P, Jacquez GM, Marcus WA (2005) Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies. *Remote Sens Environ* 95:351–367
- Goovaerts P, Albuquerque T, Antunes IMHR (2014) A spatial statistical approach for sedimentary gold exploration: a Portuguese case study. *Mathematics of planet earth*. In: *Lecture notes in earth system sciences*, pp 545–548. doi:10.1007/978-3-642-32408-6_119 (ISSN 2193-8571, ISBN 978-3-642-32407-9)
- Hale M, Plant JA (1994) Drainage geochemistry. In: Govett GJS (ed) *Handbook of Exploration Geochemistry*, vol 6. Elsevier, Amsterdam
- Harraz HZ, Hamdy MM, El-Mamoney MH (2012) Multi-element association analysis of stream sediment geochemistry data for predicting gold deposits in Barramiya gold mine, Eastern Desert, Egypt. *J Afr Earth Sci* 68:1–14
- Hronsky JMA (2004) The science of exploration targeting. In: Muhling J (ed) *SEG 2004, predictive mineral discovery under cover*. Publication 33. Centre for Global Metallogeny, University of Western Australia, pp 129–133
- Hronsky JMA, Groves DJ (2008) Science of targeting: definition, strategies, targeting and performance measurement. *Aust J Earth Sci* 55:3–12
- Geológico Instituto e Mineiro (1988) Reports from a prospecting project for tungsten, tin and associated minerals from Góis-Segura. *Metallic Mineral Prospecting Section*, Porto
- Madani AA (2011) Knowledge-driven GIS modeling technique for gold exploration, Bulghah gold mine area, Saudi Arabia. *Egypt J Remote Sens Space Sci* 14:91–97
- Maroto AG, Navarrete J, Jimenez RA (1997) Concentraciones de metales pesados en la vegetación autoctona desarrollada sobre los suelos del entorno de una mina abandonada. *Bol Geol Minero* 108–1:67–74
- McCuaig TC, Beresford S, Hronsky J (2010) Translating the mineral systems approach into an effective exploration targeting system. *Ore Geol Rev* 38:128–138

- McInnes M, Greenough JD, Fryer BJ, Wells R (2008) Trace elements in native gold by solution IPC-MS and their use in mineral exploration: a British Columbia example. *Appl Geochem* 23:1076–1085
- Mortensen JK, Chapman R, LeBarge W, Jackson L (2004) Application of placer and lode gold geochemistry to gold exploration in western Yukon. *Yukon Explor Geol* 2004:205–212
- Oliveira JT, Pereira E, Ramalho M, Antunes MT, Monteiro JH (coord) (1992) *Carta Geológica de Portugal à escala 1/500 000 (5a edição)*. Serviços geológicos de Portugal, Lisboa
- Outridge PM, Dorherty W, Gregoire DC (1998) Determination of trace elements signatures in placer gold by laser ablation—inductively coupled plasma—mass spectrometry as a potential aid for gold exploration. *J Geochem Explor* 60:229–240
- Reimann C, Filzmoser P, Garret RG, Dutter R (2008) *Statistical data analysis explained. Applied Environmental Statistics with R*, Wiley Chichester 343 pp
- Sadeghi M, Morris GA, Carranza EJM, Ladenberger A, Andersson M (2013) Rare earth element distribution and mineralization in Sweden: an application of principal component analysis to FOREGS soil geochemistry. In: Foley N, De Vivo B, Salminen R (guest eds) *Rare earth elements: the role of geology, exploration, and analytical geochemistry in ensuring diverse sources to supply and a globally sustainable resource* (special issue). *J Geochem Explor* 133:160–175
- Sadeghi M, Billay A, Carranza EJM (2014) Analysis and mapping of soil geochemical anomalies: implications for bedrock mapping and gold exploration in Giyani area. *J Geochem Explor (South Africa)*. doi:[10.1016/j.gexplo.2014.11.018](https://doi.org/10.1016/j.gexplo.2014.11.018)
- Santos Oliveira JM, Pedrosa MY, Canto Machado MJ, Rochas Silva J (1998) Impacte ambiental provocado pela actividade mineira. Caracterização da situação junto da Mina de Jales, avaliação dos riscos e medidas de reabilitação. *Actas do V Congresso Nacional de Geologia* 84/2, pp E74–E77
- Townley BK, Hérail G, Maksiav V, Palacios C, de Parseval P, Sepulveda F, Orellana R, Rivas P, Ulloa C (2003) Gold grain morphology and composition as an exploration tool: application to gold exploration in covered areas. *Geochem Explor Environ Anal* 3:29–38
- Viladevall M, Font X, Carmona JM (1999) Multidata set analysis for gold-deposit exploration criteria: application in the Catalanian Coastal Ranges (NE Spain). *J Geochem Explor* 66:183–197
- Wang CM, Cheng QM, Deng J, Xie SY (2008) Magmatic hydrothermal superlarge metallogenic systems—a case study of the Nannihu Ore Field. *J China Univ Geosci* 19:391–403
- Zhang C, Luo L, Xu W (2008) Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Sci Total Environ* 398:212–221
- Zuo R (2011) Identifying geochemical anomalies associated with Cu and Pb–Zn skarn mineralization using principal component analysis and spectrum-area fractal modeling in the Gangdese Belt, Tibet (China). *J Geochem Explor* 111:13–22