geosciences

MDPI

Article

An Application of High-Dimensional Statistics to Predictive Modeling of Grade Variability

Juri Hinz [1,*], Igor Grigoryev [2] and Alexander Novikov [1]

[1] School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123, Ultimo, NSW 2007, Australia; alex.novikov@uts.edu.au

[2] Faculty of Science and Technology, Federation University Australia, PO Box 663, Ballarat, VIC 3353, Australia; i.grigoryev@federation.edu.au

* Correspondence: juri.hinz@uts.edu.au

Received: 26 January 2020; Accepted: 19 March 2020; Published: 25 March 2020



Abstract: The economic viability of a mining project depends on its efficient exploration, which requires a prediction of worthwhile ore in a mine deposit. In this work, we apply the so-called LASSO methodology to estimate mineral concentration within unexplored areas. Our methodology outperforms traditional techniques not only in terms of logical consistency, but potentially also in costs reduction. Our approach is illustrated by a full source code listing and a detailed discussion of the advantages and limitations of our approach.

Keywords: prediction; artificial intelligence; machine learning; LASSO; cross-validation

1. Introduction

Any mining project begins with the exploration stage which includes estimation of the expected economic returns, which can be made after significant mineralization has been encountered. Naturally, each exploration stage is designed to proceed to the next stage only if the results of the current stage justify this continuation. Therefore, the key goal remains the identification of high-priority targets and the description of a conceptual model upon which the mine design and the project's net present value can be established.

There are notable commercial interests in developing methods to accelerate an exploration and to more precisely identify the most realistic economic scenario. Thereby, the focus is on making predictions for unexplored areas based on limited drilling data and using this to estimate anticipated costs and potential revenues. Our method can be very useful in this context. Namely, in situations where the data is insufficient to perform any reasonable resource estimation, an Exploration Target Statement can be obtained. However, it may be that early drilling results may delineate the Exploration Target and lead to an Inferred Mineral Resource.

In this work, we utilize recent developments in high-dimensional statistics which provide a control on prediction accuracy by the reduction of the so-called "generalization error". This error is a crucial notion in modern statistical analysis and represents a cumulative result deterioration (due to uncertainty in model specification and its estimation errors) when applying predictions to new observations. The reduction of generalization error is achieved using appropriate penalization and regularization techniques, which are usually combined with some parameter selection method. In our approach, such parameter selection is based on in-sample estimation of predictive performance, effected by the so-called "cross-validation".

This approach has been successful (see [1–3]) in estimation of very weak dependencies in large data sets. This work presents a statistically accounted decision support methodology targeted on:

- improving prediction accuracy of mineral presence, which may lead to more cost-efficient exploration,
- assessment of the spatial mineralization characteristics, which may provide an alternative viewpoint which contrasts with existing methodologies,
- introduction of advanced statistical/artificial intelligence methods into the traditional area of geostatistics.

In this work, we demonstrate how drilling data can be combined with penalized regression to establish a statistical model focused on predictive performance. The results of this approach are compared to a traditional technique.

2. Background

In the literature, the term spatial prediction refers to an interpolation by weighted averages of the observed data [4]. Thereby, any existing method used in resource definition has its own drawbacks. Prediction of grade variability is often viewed as an interpolation methodology, which has been criticized for a number of restrictive assumptions [5–7]. Traditional geostatistical methods described in [7–11] presume a certain probabilistic structure obtained from the variogram. In some sense, it is supposed that the observation of a mineral concentration at location $S \in \mathcal{O}$ follows a random field $(Z(S))_{S\in\mathcal{O}}$ indexed by points of a given geological structure $\mathcal{O} \subset \mathbb{R}^3$. Having adopted appropriate assumptions on this random field such as stationarity and its covariance structure and given measurements $(Z(S_i))_{i=1}^{I}$ at locations $S_i \in \mathcal{O}$, $i = 1, \dots I$, the estimation of $Z(S_0)$ at an unobserved point $S_0 \notin \{S_1, \dots, S_I\}$ is defined in terms of conditional expectations

$$\mathbb{E}(Z(S_0) \mid Z(S_1), \dots, Z(S_I)). \tag{1}$$

While a probabilistic framework and estimation (1) provides foundation for this work, the proposed approach is different. Instead of relying on stationarity with the inconvenient need to specify a certain dependence type, it is supposed that the dependence between spatial observations is modeled naturally, in terms of the so-called ore body—a hidden structure which contains a certain concentration of a mineral. The shape, size, and content of the ore body are random and are determined by a superposition of the so-called prime ore deposits, deposits which are modeled by ball-shape structures with uncertain radius, location and rock mass content. Having assumed that the ore deposits may occur at virtually any location of the domain, we obtain an over-determined model which can be estimated using modern techniques of high-dimensional statistics. More importantly, the proposed estimation is performed by the selection of a most appropriate model from the viewpoint of its predictive accuracy.

Remark 1. *Here, we use stimation in the statistical sense of likelihood estimation for a parameterized family of densities. In this context, the notion overdetermined stands for a situation where there are different parameter sets which represents the same likelihood maximizing density. These concepts are introduced and explained in the following sections.*

3. High-Dimensional Statistics: LASSO and Ridging

A linear regression model is characterized by the relation

$$Y = \mathbf{X}\beta^* + \varepsilon, \tag{2}$$

where

$Y = (Y_i)_{i=1}^n$ are observed response variables, $n \in \mathbb{N}$,
$\mathbf{X} = (\mathbf{X}_{i,j})_{i=1,j=1}^{n,p}$ is a *given* matrix,
$\beta^* = (\beta_j^*)_{j=1}^p \in \mathbb{R}^p$ is an *unknown* vector of coefficients,
$\varepsilon = (\varepsilon_i)_{i=1}^n$ are unobserved random variables (zero-mean, with unknown variance σ^2),
n is the sample size,
p denotes number of predictors.

Note that the matrix entries $\mathbf{X} = (\mathbf{X}_{i,j})_{i=1,j=1}^{n,p}$ are non-random by assumption. This situation is referred as fixed design. The columns $\mathbf{X}_{\cdot,j}$ for $j = 1, \ldots, p$ are called predictor realizations. The model (2) is linear because its dependence from the coefficients $\beta^* = (\beta_j^*)_{j=1}^p$, is linear. As mentioned above, the predictor (p variables) can be obtained via non-linear transformation of some (multivariate) explanatory factors. Similarly, the response variable can also be described by some non-linear transformation of the actual observations, for instance by their measurement on a different (logarithmic) scale. On this account, linear models are able to capture non-linear dependencies between random phenomena. In applications of linear models, the estimation of β^* from observation of $\mathbf{X}$ and Y is typically unstable when columns of $\mathbf{X}$ are correlated. In this case, the least-squares estimation may be ill-posed (if eigenvalues in the matrix $\mathbf{X}^\top\mathbf{X}$ are small). Furthermore, there are applications with more potential explanatory variables than observations ($n < p$). In such situations, the ridge regression [12] (p. 215) which penalizes the size of the regression coefficients, suggests solving a penalized least squares problem

$$\hat{\beta}(\lambda) = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \left(\|Y - \mathbf{X}\beta\|^2 + \lambda \|\beta\|^2 \right) \tag{3}$$

for Euclidean norm $\|\cdot\|$, $\lambda > 0$ with an appropriate penalization parameter. Note that even in the case that $\mathbf{X}$ does not have full rank, this problem possesses a unique solution, which is obtained from the equation

$$(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{1})\hat{\beta}(\lambda) = \mathbf{X}^\top Y,$$

where $\mathbf{1}$ stands for the identical matrix. The ridge penalty λ causes a shrinkage of the coefficients

$$\hat{\beta}(\lambda) \to 0 \quad \text{for } \lambda \to \infty.$$

This usually yields a variance reduction

$$\operatorname{Var}(\hat{\beta}(\lambda)) < \operatorname{Var}(\hat{\beta}(\lambda')) \quad \text{for } \lambda > \lambda' > 0$$

but introduces a systematic error (bias)

$$\mathbb{E}(\hat{\beta}(\lambda)) \neq \beta^*, \quad \lambda > 0.$$

The LASSO is introduced in [1] as a regularization method for a linear regression which uses an ℓ_1 penalty to achieve a sparse solution and the name "LASSO" was also introduced as an acronym for Least Absolute Selection and Shrinkage. With the goal of the variable selection in mind, the LASSO brings a major improvement from the viewpoint of the generalization error reduction since LASSO determines $\widehat{\beta}(\lambda)$ as a minimizer

$$\hat{\beta}(\lambda) = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \left(\|Y - \mathbf{X}\beta\|^2 + \underbrace{\lambda \|\beta\|_1}_{\text{penalization}} \right), \tag{4}$$

with l_1-norm (instead of the Euclidean norm) on the coefficients

$$\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|, \qquad \beta = (\beta_j)_{j=1}^{p} \in \mathbb{R}^p. \tag{5}$$

Note that in difference to the ridge regression, this minimization is not solvable by simple linear algebra but needs quadratic programming or related algorithms. More importantly, such a solution is appealing because usually many coefficients in $\hat{\beta}(\lambda)$ vanish (which yields a kind of model selection by variables reduction). Let us explain why $\hat{\beta}(\lambda)$ is usually a sparse vector. Finding the unconstrained minimum

$$\beta \mapsto \|Y - \mathbf{X}\beta\|^2 + \lambda\|\beta\|_1$$

is equivalent to finding the constraint minimum

$$\beta \mapsto \|Y - \mathbf{X}\beta\|^2 \quad \text{subject to} \quad \|\beta\|_1 \leq R$$

for an appropriate boundary parameter $R = R(\lambda)$. Since the ball $\|\beta\|_1 \leq R$ has sharp edges, the solution $\widehat{\beta}(\lambda)$ is a sparse vector.

Using advanced numerical schemes, it is possible to determine $\widehat{\beta}(\lambda)$ for many (all relevant) penalization parameters $\lambda \in \mathbb{R}_+$, thus one can chose the most appropriate LASSO penalization $\widehat{\lambda}$ through the so-called cross-validation (CV). Here are typical steps for the LASSO:

1. Divide all observations into train and test sets (i.g., 90 % train, 10 % test). Do this randomly i.g., in 10 different ways.
2. Determine $\widehat{\beta}(\lambda)$ on the train data, and with this $\widehat{\beta}(\lambda)$ calculate the averaged mean-square error on the test data. Average these errors over all 10 samples. Do this for (all relevant) $\lambda \in \mathbb{R}_+$.
3. Determine a penalty parameter $\widehat{\lambda}$, which yields the best averaged result on the test data.
4. Infer the model coefficients as $\widehat{\beta}(\widehat{\lambda})$.

The literature suggests that various types of optimization problems can be solved with the ℓ_1-norm as a regularizer. An overview of some basic optimization algorithms for convex problems, with an emphasis on aspects of particular relevance to regularized estimators such as the LASSO can be found in [3].

4. Proposed Design

Suppose that within a predetermined mineralized rock mass $\mathcal{O} \subset \mathbb{R}^3$ there are sparsely distributed observations at locations which are based on samples obtained from drill-core assay. Then the distribution is being estimated at each spatial point within the targeted range. To address such problem, a number of techniques have been suggested which are usually based on interpolation, such as inverse distance-weighted and radial basis functions, global polynomial and local polynomial interpolations (see [13]). However, our approach is different and originates from the observation that the occurrence of worthwhile ores is associated with an ore body, which is a geological structure containing a distinct (compared to the average area) concentration of metals. Specifically, it is proposed to estimate ore reserves through the reconstruction of such an ore body, whose shape and distribution should give a best possible explanation of the taken samples. The core idea is that having determined a notably high concentration $y \in \mathbb{R}_+$ at a location $x \in \mathcal{O} \subset \mathbb{R}^3$, the conditioned probability density for the mineral concentration at any other point $x' \in \mathcal{O} \subset \mathbb{R}^3$ depends on the Euclidean distance $\|x - x'\|$ to the location and on the concentration measured there. Although this dependence is difficult to quantify, it is assumed that

the higher y and the smaller the distance $\|x - x'\|$, the more likely it is to find the mineral (in high concentration) at the point x'. This idea is represented in terms of a kernel density

$$k : \mathbb{R}^3 \to \mathbb{R}_+, \qquad x \mapsto k(x), \qquad \int_{\mathbb{R}^3} k(x)dx = 1$$

with the assumption that given a location $l \in \mathcal{O}$, the function $y \mapsto k(x - l)$ represents a a spherical ore body, concentrated around a point $l \in \mathcal{O}$. This statement can be quantified as follows: given the occurrence of a prime ore deposit located at point $l \in \mathcal{O}$, the expectation of the (averaged) gold concentration within each rock mass volume $V \subset \mathcal{O}$ is represented by a random variable whose (conditioned) expectation is given by

$$\int_V k(v - l)dv.$$

To reflect the expected concentration decay depending only on the distance, without taking into account any particular direction, it is assumed that

$$k(x) \quad \text{depends on Euclidean distance } \|x\| \text{ only and decreases with } \|x\|.$$

Suppose further that such mineral deposits can be located at $p \in \mathbb{N}$ different points $(l_j)_{j=1}^p \subset \mathcal{O}$ and have different intensities $(\beta_j)_{j=1}^p$. Note that the intensity coefficient β_j describes the ore concentration and, according to our modeling, the dependence of concentration on the coefficients $(\beta_j)_{j=1}^p$ is linear.

To resemble an ore body by a superposition of such deposits, we suppose that the expectation of the (averaged) mineral concentration within each rock mass volume $V \subset \mathcal{O}$ is represented by a random variable whose (conditioned) expectation is given by

$$\sum_{j=1}^{p} \beta_j \int_V k(x - l_j)dx.$$

Finally, we agree to explicitly construct such random variables for each rock mass volume $V \subset \mathcal{O}$ in terms of

$$Y(V) = \sum_{j=1}^{p} \beta_i \int_V k(x - l_j)dx + \frac{\sigma}{\eta(V)}\varepsilon(V), \quad V \in \mathcal{B}(\mathcal{O}) \tag{6}$$

indexed by all Borel sets $\mathcal{B}(\mathcal{O})$ of $\mathcal{O}$. Here, $\sigma \in \mathbb{R}_+$ is a pre-specified parameter and the family $\{\varepsilon(V) : V \in \mathcal{B}(\mathcal{O})\}$ represents a so-called Gaussian white noise. This is a family of Gaussian random variables which satisfy

$$\varepsilon(V) \sim N(0, \eta(V)), \qquad \mathrm{Cov}(\varepsilon(V), \varepsilon(V')) = \eta(V \cap V'), \quad V, V' \in \mathcal{B}(\mathcal{O}) \tag{7}$$

where η denotes the Lebesgue measure and N stands for normal distribution.

Having agreed on the assumption (6) and (7), determining the shape of the ore body boils down to the estimation of the parameters $l = (l_j)_{j=1}^p$ and $\beta = (\beta_j)_{j=1}^p$, given the observations $y = (y_i)_{i=1}^n$ of the mineral concentration within a number $n \in \mathbb{N}$ of volume samples $(V_i)_{i=1}^n$. A generic method therefore is the maximization of the log-likelihood density

$$L_{l,\beta}(y) = \log\left(\Pi_{i=1}^n \exp(-\frac{\|y_i - \mu_i(\beta, l)\|^2}{2\sigma^2\eta(V_i)^{-1}})/\sqrt{2\pi\sigma^2\eta(V_i)^{-1}}\right)$$

where the numbers $(\mu_i(\beta, l))_{i=1}^n$ stand for the expectation of the mineral concentration within the sample volume V_i under the assumption that the deposit locations are $l = (l_j)_{j=1}^p$ and their intensities are $\beta = (\beta_j)_{j=1}^p$:

$$\mu_i(\beta, l) = \sum_{j=1}^{p} \beta_j \int_{V_i} k(x - l_j)dx, \qquad i = 1, \ldots, n.$$

Following the above maximum-likelihood principle, a maximum of the function

$$(\mathbb{R}^3)^p \times \mathbb{R}^p \to \mathbb{R}, \quad (l, \beta) \mapsto L_{l,\beta}(y)$$

must be determined. However, a straight-forward application of such a concept has a number of drawbacks:

- The log-likelihood function is not convex, and the dimension $4p$ of its arguments is in general very high.
- There are numerical issues determining its global maximum.
- The choice of the number $p \in \mathbb{N}$ of prime ore deposits is arbitrary and there is no obvious way to determine this number. Choosing p too low results in a poor model, whereas if p is too high, then a large number of parameters may cause over-fitting and deteriorate the predictive performance of the model.

In what follows, a different approach to prediction is suggested. The idea is based on advances in high-dimensional statistics. Utilizing these tools, the obstacles described above can be overridden. From an abstract perspective, a moderate-dimensional non-convex problem is reformulated into an extremely high-dimensional but convex optimization problem. Furthermore, a built-in penalization of parameters with the focus on the minimization of the generalization errors leads to dimension reduction while maximizing the model's predictive performance.

Let us explain our approach. Having supposed that the volume samples $(V_i)_{i=1}^n$ have equal sizes

$$\eta(V_i) = \nu_0 \in \mathbb{R}_+ \quad i = 1, \ldots, n$$

and assuming that the locations $(l_j)_{j=1}^p$ of all prime ore deposits are fixed, we observe that the log-likelihood function must be maximized with respect to the intensity parameters $\beta = (\beta_j)_{j=1}^p$. However, the problem of finding a maximum of

$$\mathbb{R}^p \to \mathbb{R}, \quad \beta \mapsto L_{l,\beta}(y)$$

can be formulated using a standard framework of linear models, where the unknown coefficients $\beta \in \mathbb{R}^p$ are estimated under the assumption that $Y = (y_i)_{i=1}^n$ are observations of the response variable $Y = \mathbf{X}\beta + \varepsilon$ with independent identically normally distributed random variables $\varepsilon = (\varepsilon_i)_{i=1}^n$ and a non-random design matrix

$$\mathbf{X} = (\mathbf{X}_{i,j})_{i,j=1}^p, \quad \mathbf{X}_{i,j} = \int_{V_i} k(x - l_j)dx, \qquad i = 1, \ldots, n, \;\; j = 1, \ldots, p.$$

Note, however that the centers $l = (l_j)_{j=1}^p$ of prime ore deposits are fixed. To address this issue, we propose that their total number p is sufficiently high, which allows a potential prime ore deposit to exist at virtually any location. At the same time, we suggest to use a penalized regression. Instead of solving the ordinary regression

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \|Y - \mathbf{X}\beta\|,$$

we address an adequate formulation in the spirit of LASSO (4) with l_1-norm (5) on the coefficients.

As mentioned above, the penalized regression has several advantages: It allows a view from the perspective of high-dimensional statistics whose methods are designed to solve linear models under the assumption that the number of explanatory variables exceeds (by magnitudes) the number of observations, having in mind that only a few (not knowing which) of the explanatory variables have non-zero coefficients. Furthermore, the penalized regression in the LASSO formulation performs an automated variable selection, by determining those non-zero coefficients which maximize the predictive performance through an estimation of the generalization error via cross-validation. Note that this method is appropriate for our statistical estimation since we suppose that at virtually each location there could be a prime ore deposit of unknown intensity. Realistically, almost all intensities shall vanish, but it is not known which deposits contribute with non-zero intensity.

Finally, let us emphasize that the goal of a geostatistical model is not to fit/smooth a given sample of observations, but to infer a mineral concentration in the entire domain with enhanced predictive ability. Of course, all relevant physical and geological considerations (for instance, the mineral clustering described in terms of the ore body concept) must be incorporated in some way into such a model. These goals are approached by our methodology.

5. Experiment

We illustrate our techniques on a data set which, in a real-world situation, would represent commercial in confidence information about the geological situation. This is to avoid problems with information disclosure which we outline next. Typically, in early stages of exploration, the investors have limited access to data necessary to make an investment decision. This means that any leak, reference or inaccurate statement (pessimistic or over-optimistic) may lead to a change of share prices and/or give rise to insider trading allegations if new information was not properly disclosed. That is, exploration companies tend to operate privately and competitively, which explains the difficulties in obtaining, analyzing, and presenting large and interesting real-world data sets to the public. For this reason, the following study may appear limited from many perspectives as it does not use additional geological information which would normally be available in a real-world exploration project. The reader should consider the following section as a conceptual illustration of the methodology, rather than a real-world geological investigation. We illustrate our techniques by a data set which represents restricted information about geological situation, in order to avoid problems with information disclosure

5.1. Experimental Data Presentation

A dataset was obtained from a mine site at the production phase during the grade control program. Geologically, it relates to porphyry deposits and represents an important resource and the dominant source of copper and gold. It is located approximately 100 km north of Orange NSW and contains two major elements as a source of revenue: copper (Cu) and gold (Au) with triuranium octoxide (U_3O_8) as sub-product.

Remark 2. *Note that since only gold data was available, for the purpose of demonstration of the procedures described in Section 4 it is assumed that only gold is of primary interest. Lithology, stratigraphy and lithochemistry have not been available and hence have not been considered.*

Assume the pit carried out a grade control exercise using samples obtained from vertical data, in order to delineate the gold-bearing ore for the next four benches. Thereby, 119 vertical blastholes horizontally spaced as 5×2.5 m were developed from the surface by rotary drilling practice as shown in Figure 1 and

sampled in 2.5 m vertical intervals. Basic geometrical and statistical descriptions for the experimental data can be found in Tables 1 and 2. Geometrical limits of the blasthole drilling grid can be found in Table 1.

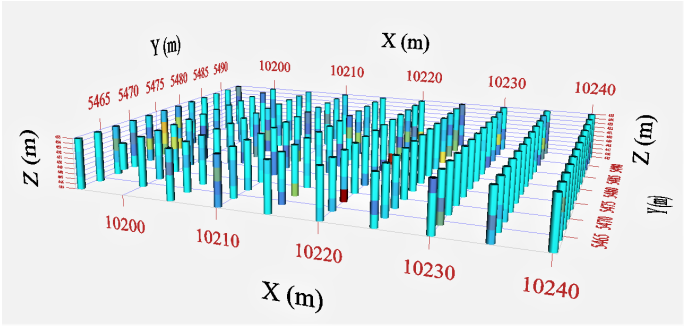


Figure 1. Rectangular blasthole pattern. Gold content is shown in the solid-trace style sized according to gold grade, high grades are marked as red.

On-site analysis revealed the presence of multiple clusters of gold close to the vertical blast holes (10 m each). No accurate assumptions regarding gold distribution were made at the time of the grade control program due to the lack of reliable data in between the 2.5 m intervals. The existence of gold within the bench (the area is 245 m×33 m, or 8085 m^2) provided an opportunity for a test of the proposed methodology for estimation of the distribution on un-sampled locations. Using actual sampling data (Figure 1) and Tables 1 and 2, Cartesian coordinates and grades were assigned to 5975 valid Au samples. Considering that exploratory drilling cannot meaningfully take place, it was revealed that

1. Gold content is represented by several, potentially narrow structures.
2. The edges of the structures are poorly defined, making their definition at the time of extraction difficult.
3. Economic cut-off grade is unknown. Lowering the cut-off grade may result in a steady increase in the tonnage to be extracted and milled. If the cut-off grade increased, the impact of dilution may be large.

Knowledge of the content between the sampled locations is highly desirable because it helps in making gold projections from one horizontal section to another and thereby for identifying high-priority structures (extraction targets) before the actual exploration commences. Thereby a typical dilemma lies in predicting the distribution between sampled locations

Table 1. Bench blasthole geometry.

	X Min	X Max	Y Min	Y Max	Z Min	Z Max	Block
Geometry	10,195 m	10,240 m	5461 m	5494 m	445 m	455 m	25,523 m^3

Table 2. Statistics reporting on numerical variables by quantiles in gold g/t.

Total	Min	Max	μ	SD	σ^2	Q:0.05	Q:0.10	Q:0.25	Q:0.75	Q:0.95
5975	0.01	19.22	1.97	2.95	8.7 g/t^2	0.04	0.07	0.23	2.48	8.73

The area has been split into four horizontal sections to depict the third dimension (depth) in order to explore the distribution. By placing the so-called isograms of content intervals, the gold grades can be interpolated (g/tonne) to arbitrary (documented) points. Figures A1–A4 illustrate the result obtained from kriging.

5.2. Method Implementation

The input data for modelling are as follows: 5975 valid samples, blasthole ID, $X; Y; Z$ coordinates, the positions of drilling collars on surface, the length of the sampled sections of each blasthole (given in meters), the true depth (given in metres) and the gold grade (given in g/tonne). The estimation of model parameters through an iterative application of LASSO methodology has been implemented (see Appendix B) as follows

(a) At the beginning a number ($p = 700$) of deposit centers is randomly scattered around the area to be investigated. Each kernel is represented by a normal density whose mean equals the corresponding deposit center, with the same variance.
(b) The coefficients are estimated using standard LASSO method with 10-fold cross validation for penalization parameter selection.
(c) Each deposit center, which has non-zero coefficients, is disturbed (its kernel variance and kernel center are changed randomly). The deposit with zero coefficients are removed and those obtained via disturbance are added to the potential set of deposits.
(d) The steps (b) and (c) are repeated (five to seven times).

6. Results

The distribution of gold has been successfully estimated by solving a high-dimensional convex optimization problem using standard techniques, returning results which are potentially applicable for evaluation. Both procedures (LASSO and kriging) indicate the presence of a trend of increasing grade with depth in the NE direction and the existence of internal dilution in SE direction. Both methods suggest the existence of localized occurrence of middle/high localized grades within the underlying subsection of the mining bench (flitch), the development of which may be economically profitable. Our methodology uses standard open-source implementations, is stable and requires low computing resources. More importantly it is fast: the average run time on a standard PC ranges from 110 s to 160 s. One crucial aspect of our technique is that the algorithm determines virtually the same solution for a wide range when used with different starting values and model parameters, which is in line with our expectations. Note that, in contrast with traditional interpolation, the estimated mineral concentration is smooth and plane (in the sense that it exhibits moderate spatial changes and extreme values). A comparison of LASSO computations with those obtained from kriging provides clear evidence that both methodologies yield similar results. However, the models are conceptually different: unlike kriging, the LASSO does not require variography and model fitting. Since both methods are based on different concepts, that the analysis of prediction disagreements may be helpful in uncovering potential shortcomings in both methods.

The distribution of gold has been successfully estimated by solving a high-dimensional convex optimization problem using standard techniques, returning results which are potentially applicable for evaluation. Both procedures (LASSO and kriging) indicate the presence of a trend towards North-East direction with overall increase of grade with depth and the existence of internal dilution in SE direction. Both methods suggest the existence of localized occurrence of middle/high localized grades within the underlying flitch, the development of which can be economically profitable. Our methodology uses standard open-source implementations, is stable, and requires low computing resources.

More importantly, it is fast: the average run time on a standard PC ranges from 110 s to 160 s. One crucial aspect of our technique is that the algorithm determines virtually the same solution for a wide range with different starting values and model parameters, which is in line with our expectations. Note that in difference to traditional interpolation, the estimated mineral concentration is smooth and plane.

We have concluded that the generalization error control avoids undesirable concentration spikes and turbulences inherent to interpolation methods. A comparison of LASSO computations with those obtained from kriging provides clear evidence that both methodologies render semantically comparable plots. However, the models are conceptually different: unlike kriging, the LASSO does not require variography and model fitting. Since both methods are based on different concepts, the comparison of prediction disagreements can be helpful in uncovering the singularities and shortcomings of both methods.

Interpretation

Kriging: To explore and compare the distributions, four gold favourability maps of 2.5 m thick layers were created using kriging. The cut-off grade = 0.5 g/t has been artificially applied to all maps to display the contrast between low and higher grades. The realizations (Appendix A) shown in Figures A2 and A3 indicate that depending on the depth, the bench (flitch) 10 m may contain gold concentrations $\geq$19.00 ppm or 19 *Au* g/t. Low *Au* grades are present at the locality 5470 N/10,230 E. Relatively high grades may occur at 5485 N/10,210 E. The cross-sections (Figures A3 and A4) display the presence of developing trend in increasing grades with depth. The extraction and processing of rock mass of underlying flitch can be economic. However, the Eastern region shows a very low reading. From an engineering point of view, the depth below RL445 should be worthy for close attention because the underlying horizontal levels are likely to have larger amount of gold. Comparing four cross-sections side-by-side (Figures A1–A4), one may conclude that the distribution is not gradational. Obviously, the sampling intervals 2.5 m are too large to make an accurate assumption on gold potential.

LASSO: At 1.875 m depth the model predicts the presence of a developing trend towards North-East direction with overall increase of grade and the existence of internal dilution in South-East and East directions (Figure 2). At 3.75m depth a stable trend identified and model indicates significant geometrical changes accompanied by the mean and median displacements (Figure 3). At 5.625 m model renders consistent increase of the apparent grade with the depth. The distribution reflects insignificant reshaping and suggests the displacement of maximums (Figure 4). Finally, at 7.5 m depth (Figure 5) the apparent grade increases considerably with the increase of depth. Contours illustrate a preferential trend in the northeasterly direction. The extraction of the underlying level, i.e., below RL445 can be economically profitable. The adjacent flitch at the Easterly locality requires for careful grade control.

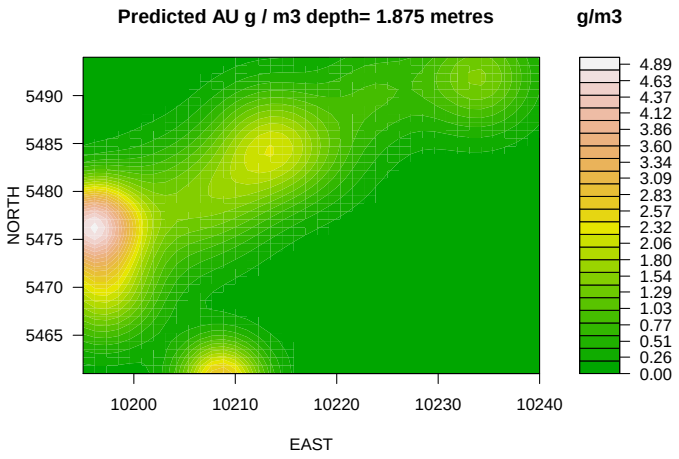


Figure 2. Predicted production target favorability map in Au g/m^3 at depth = 1.875 m, where high grades contrasting with low grades. Model indicates the presence of a developing trend towards North-East direction with overall increase of grade and the existence of internal dilution in SE and E directions assigned the color green.

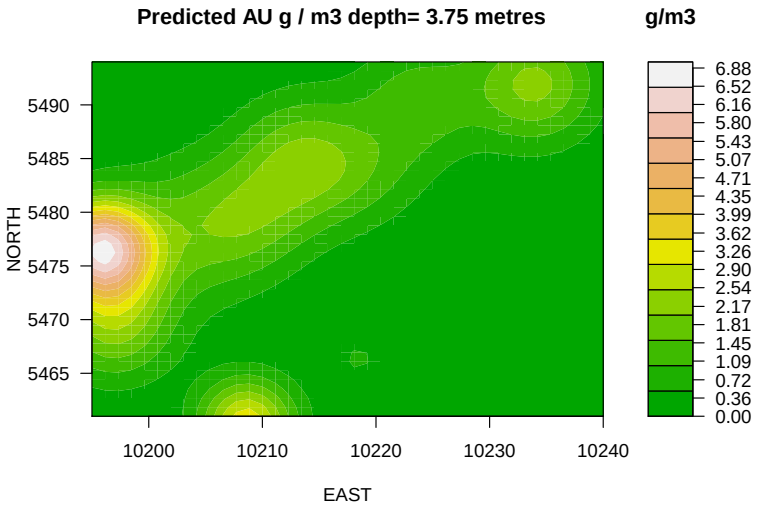


Figure 3. Production target favorability map at depth of 3.75 m in Au g/m^3. Stable trend identified, the modeled indicates significant geometrical changes accompanied by the mean and median displacements.

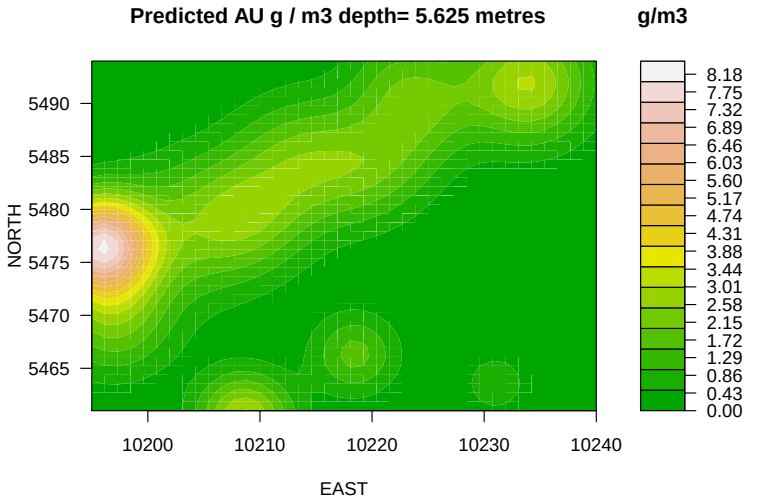


Figure 4. Predicted gold content at 5.625 m below the surface. Model renders a consistent increase of the apparent grade with the depth. The distribution reflects insignificant reshaping and suggests the displacement of maximums.

Comparison: Note that the above methods are not competing as they investigate the same problem from different perspectives using different concepts and metrics. However, both methods indicate the presence of an increasing trend in grades with depth. From an engineering perspective, the rate of change of gold grades in the "kriged" models is to0 abrupt, in particular between RL455–RL447.5 (as depicted in Figures A2–A4). Furthermore, the borders of the extraction targets are not well defined. The shape of dilution (waste or barren rock) is more critical during plant operation than the location of the erratic and insignificant concentrations of the gold. The *Cu/Au* deposits such as Marvin are usually extracted by blocks, in a 10 m lift by a shovel with bucket capacity $>$20 m^3. That means the impact of the excessive amount of dilution may be extremely large. Although little evidence is available from the data to suggest that LASSO prevails over kriging in terms of predictive accuracy, note LASSO technique is designed to select a model with the strongest predictive performance—a quality target, which is absolutely not present in the kriging approach.

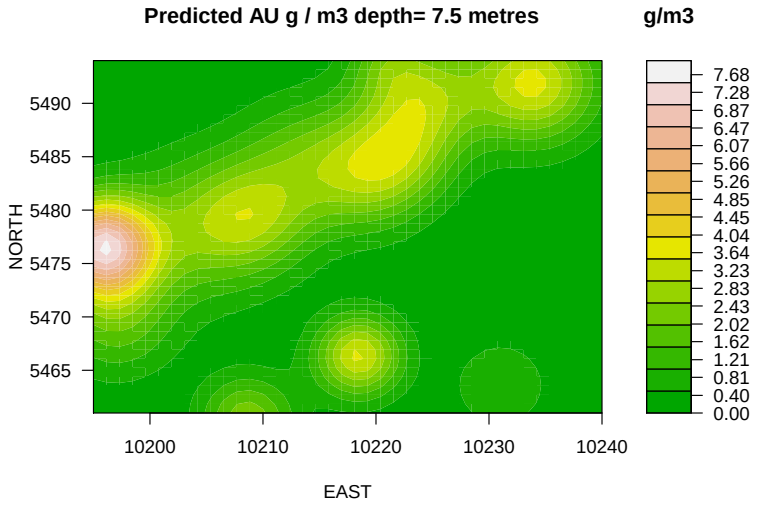


Figure 5. A realization of cross-section at depth of 7.5 m. The apparent *Au* grade increases considerably with the increase of depth. Contours illustrate a preferential trend in the northeasterly direction. The extraction of the underlying level, i.e., below RL445 can be economically profitable. The adjacent flitch at the Easterly locality requires careful grade control.

Advantages and limitations: Being designed to deal with very large data sets, out methodology may be applicable to a huge amount of geological data: in a test study, the present code was able to handle more than 50.000 samples. Furthermore, for Reverse Circulation (RC) samples (which are usually large, often longer than 3.0 m and much less variable) the analysis by kriging is complex so that LASSO provides solutions significantly faster. Moreover, the variography and model fitting are not required for LASSO, unlike for kriging which is based on anisotropy depicted by variogram in different directions. That is, in cases when an ore body is unstructured and the shapes of low-grade boundaries are gradational, LASSO can assist in fast decision making on equipment. Finally, let us mention the advantages implementation: LASSO implementation as R-package *glmnet* is a part of open-source environment and benefits from numerous data analysis and visualization techniques which are free, well-maintained, and easily accessible. However, this advantage comes at the cost of dealing with advanced scientific computing methods which require appropriate skills.

Outlook: The authors conclude that a further development of their methodology can yield a user-friendly implementation within R-environment, similarly to our code listed in Appendix B, whose use merely requires regular drilling data updates. Such implementation can be easily integrated into any on-board drilling software. Using this, the exploration and extraction crews can observe in real time the effects of their decisions in the search for high-priority mineralization.

Author Contributions: Conceptualization, methodology, implementation, validation, formal analysis, visualization, J.H., A.N., and I.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: The authors are grateful to two anonymous reviewers for their careful reading of the manuscript and insightful remarks.

Appendix A

The estimate of the distribution for the gold-bearing ores within Marvin obtained from kriging. The locations of blasthole collars are marked by crosses.

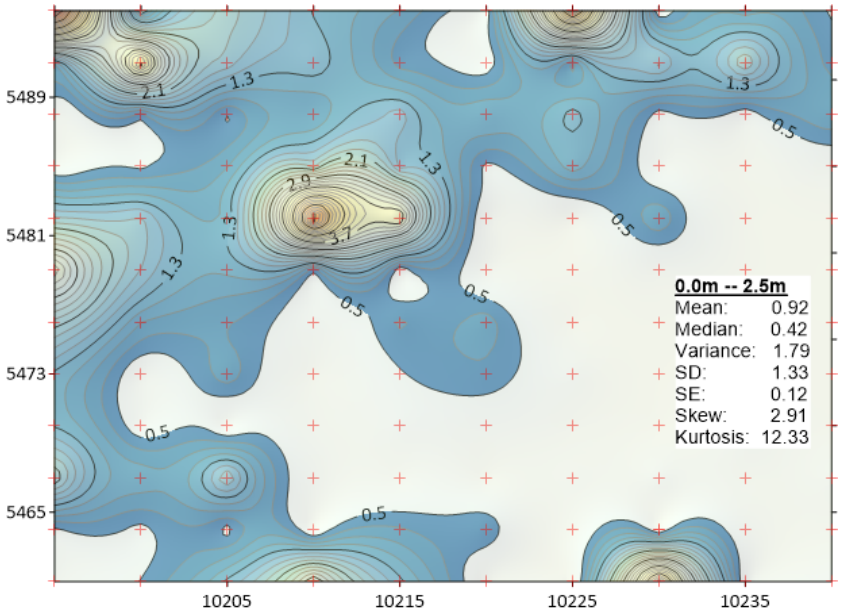


Figure A1. Distribution of ore in the form of isopath from RL455 to RL452.5 obtained from kriging. RC sampling interval (2.5 m length) from 0.0 m–2.5 m., blasthole collars are marked as red crosses.

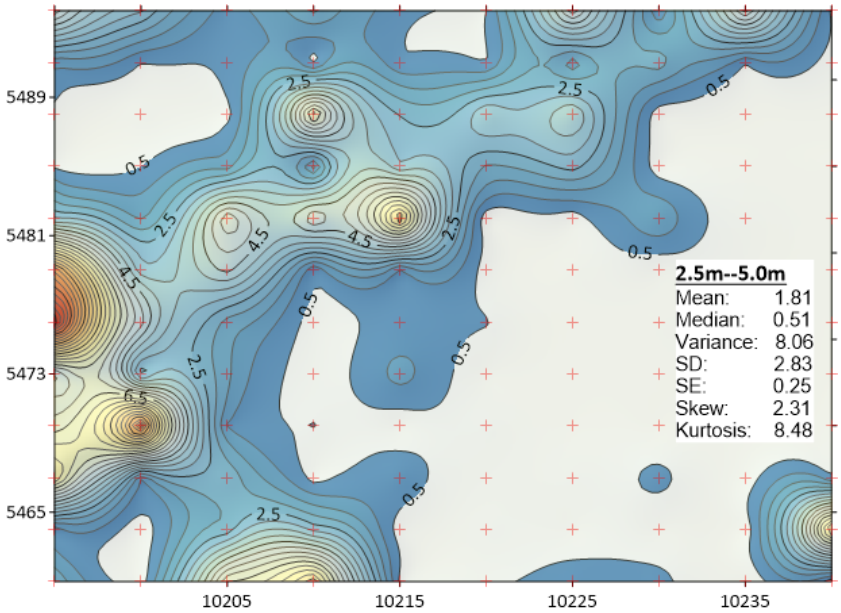


Figure A2. Distribution of ore from RL452.5 to RL450, kriging, sampling interval 2.5 m–5.0 m. The trend in higher grades is evident.

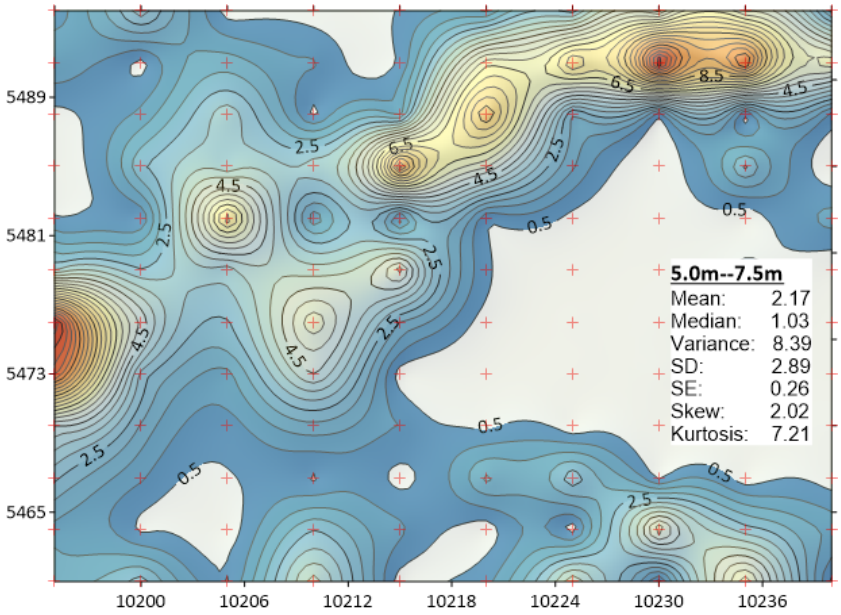


Figure A3. Distribution of ore from RL450 to RL447.5, kriging, sampling interval 5.0–7.5 m.

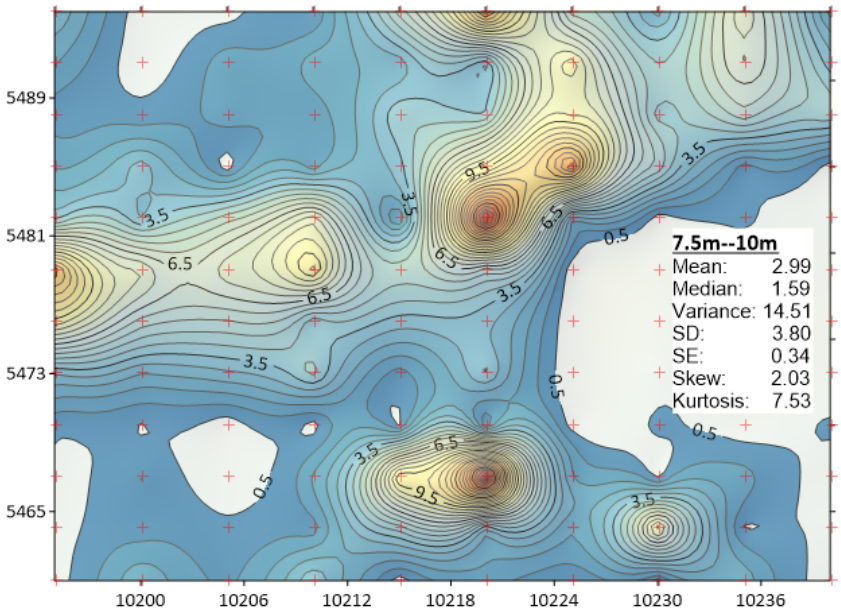


Figure A4. Distribution of ore from RL447.5 to RL445, sampling interval 7.5–10 m, consistent increase of grades with the depth.

Appendix B

This section illustrates a code implementing our techniques for Marvin data using R.

```
rm(list=ls(all=TRUE))   # clean all data
#install.packages("mvtnorm") # install packages
#install.packages("glmnet")  # if necessary
require(mvtnorm) # load packages
require(glmnet)
#############################################################################
# set your working directory
setwd("/data/windows/Documents/Codes/Rcodes/")
#
set.seed(123)
#############################################################################
dat<-read.table("Marvin1.csv") # load data from file
dat<-matrix(data=dat$V1, ncol=7, byrow=TRUE) # process data
dat<-matrix(data=as.numeric(dat[, -1]), ncol=6, byrow=FALSE)
dat<-dat[, -3]
colnames(dat)<-(c("x", "y", "z1", "z2", "au")) # the observations consist of
#  x, y, z1  beginning of segment
#  x, y, z2  end of segment
# au  gold concentration
#############################################################################
resp<-dat[, 5] # take gold concentration as response variable
obs<-dat[, 1:4]
#############################################################################
centers<-matrix(data=0, nrow=700, ncol=3)     # data for ore bodies
# first three columns for mean
# other nine for variance matrix
sigmas<-rep(1, nrow(centers))
#############################################################################
# create potential ore bodies randomly
#############################################################################
centers[, 1]<-runif(n=nrow(centers), min=min(obs[,1]), max=max(obs[, 1]))
centers[, 2]<-runif(n=nrow(centers), min=min(obs[,2]), max=max(obs[, 2]))
centers[, 3]<-runif(n=nrow(centers), min=min(obs[,3:4]), max=max(obs[, 3:4]))
#
sigmas<-runif(n=length(sigmas), min=3, max=3)
#
#############################################################################
make_model_matrix<-function(obs, centers, sigmas)
{
#############################################################################
# create model matrix
#############################################################################
# each segment contains gold concentration derived from random normal
# density originated from ore body
# rows stand for observations (gold concentration in segments)
# columns stand for ore bodies
# explanatory coefficient are obained from density integral along each segment
# which is calculated by trapezoidal rule
```

```
#########################################################################
mod_mat<-matrix(data=0, nrow=nrow(obs), ncol=nrow(centers))
for (j in 1:ncol(mod_mat)) {
center<-matrix(centers[j, 1:3], byrow=TRUE, nrow=nrow(obs), ncol=3)
p1<-obs[, c(1,2,3)]-center
p2<-obs[, c(1,2,4)]-center
mod_mat[,j]<-(apply(FUN=dmvnorm, X=p1,MARGIN=1,sigma=diag(nrow=3,ncol=3,x=sigmas[j]^2))+
apply(FUN=dmvnorm, X=p2, MARGIN = 1, sigma=diag(nrow=3, ncol=3, x=sigmas[j]^2) ) )/2  }
return(mod_mat) }
#
library(glmnet) # load library
#
#########################################################################
#  BEGIN RUN THIS SEVERAL TIMES
#########################################################################
#
max_iter<-7
#
for (k in 1:max_iter)   {
if (k>1)
{
sigmas<-sigmas_new
centers<-centers_new

if (k<max_iter-3) {

sigmas<-c(sigmas, sigmas*runif(n=length(sigmas), min=0.9, max=1.1))

centers<-rbind(centers, centers+ cbind(
runif(n=nrow(centers), min=-1, max=1),
runif(n=nrow(centers), min=-1, max=1),
runif(n=nrow(centers), min=-1, max=1) )  ) } }
#########################################################################
#
# Create model matrix
#
#########################################################################
model_mat<-make_model_matrix(obs=obs, centers=centers, sigmas=sigmas)
#########################################################################
#
#  Make explanatory matrix
#
#########################################################################
EXPL<-cbind(1, model_mat)   # define model matrix
colnames(EXPL)<-c( "i0", paste("C", rep(1:ncol(model_mat)), sep=""))
index<-1:ncol(EXPL)-1
names(index)<-colnames(EXPL)
#########################################################################
#
#  Make glm fit
#
#########################################################################
#
cvobj<-cv.glmnet(x=EXPL,y=resp, intercept=FALSE) # try with gaussian model
#########################################################################
#
# Extract non-zero coefficients
#
#########################################################################
co<-coef(cvobj)     # most of the coefficients are zero
non_zero<-which(co!=0)
non_zero_names<-names(co[non_zero,])
data_index<-index[non_zero_names]
sigmas_new<-sigmas[data_index]
centers_new<-centers[data_index,]
#########################################################################
plot(centers_new[,1], centers_new[,2])
print(sigmas_new)
print(non_zero_names)
#########################################################################
# Watch the animation run
#########################################################################
}
betas<-co[names(co[,1 ])!="(Intercept)", 1]

```

```
zlevels<-seq(from=min(dat[,3]), to=max(dat[, 3]), length=5)
res_dim1<-40
res_dim2<-40

xgrid<-seq(from=min(dat[,1]), to=max(dat[, 1]), length=res_dim1)
ygrid<-seq(from=min(dat[,2]), to=max(dat[, 2]), length=res_dim2)

for (i in 1:length(zlevels))
{
L<-cbind(expand.grid(X=xgrid, Y=ygrid), zlevels[2], zlevels[i] )
M<-cbind(1, make_model_matrix(obs=L, centers=centers, sigmas=sigmas))

zgrid<-matrix(nrow=res_dim1, ncol=res_dim2, data=M%*%betas, byrow=FALSE)

pdf(paste("/Igor/data/windows/Documents/Codes/Rcodes/
", i, ".pdf", sep="" ) , width=7, height=5, paper='special')
filled.contour(x = xgrid,y = ygrid,
z = zgrid, color.palette = terrain.colors,
plot.title = title(main =paste("gold concentration at depth",zlevels[i],"m"),
xlab = "meters x-direction",ylab = "meters y-direction"),
key.title = title(main="gold \n  g/m3"),
key.axes = axis(4, round(seq(from=min(zgrid), to=max(zgrid), length=20),  digits = 2) )
)
dev.off()
filled.contour(x = xgrid,y = ygrid,
z = zgrid, color.palette = terrain.colors,
plot.title = title(main = paste("Predicted AU g / m3 depth=",
zlevels[i], "metres", cex.main=""),
xlab = "EAST",ylab = "NORTH"),
key.title = title(main="g/m3"),
key.axes = axis(4, round(seq(from=min(zgrid),
to=max(zgrid), length=20), digits = 2) )) }
```

References

1. Tibshirani, R. Regression and Shrinkage via LASSO. *J. R. Stat. Soc. B* **2007**, *58*, 267–288.
2. Tibshirani, R.; Bien, J.; Friedman, J.; Hastie, T.; Simon, N.; Taylor, J. Strong Rules for Discarding Predictors in Lasso-type Problems. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2012**, *74*, 245–266. [CrossRef] [PubMed]
3. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.S. Statistical Learning with Sparsity the Lasso and Generalizations. In *Monographs on Statistics and Applied Probability*; CRC Press: Boca Raton, FL, USA, 2015.
4. Webster, R.; Oliver, M. *Geostatistics for Environmental Scientists*, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 2007.
5. Shurtz, R.F. A critique of A Journel's "The deterministic side of geostatistics". *Math. Geol.* **1985**, *17*, 861–868. [CrossRef]
6. Philip, G.M.; Watson, D.F. Geostatistics and spatial data. *Mathe. Geol.* **1986**, *18*, 505–508. [CrossRef]
7. Isaaks, E.H.; Srivastava, R.M. *Applied Geostatistics*; Oxford University Press: Oxford, UK, 1989; pp. 131–916.
8. David, M. *Geostatistical Ore Reserve Estimation*; Elsevier Scientific Pub. Co.: Amsterdam, The Netherlands, 1977.
9. Hengl, T. *Practical Guide to Geostatistical Mapping of Environmental Variables*; Manning Publications: Shelter Island, NY, USA, 2007.
10. Li, J.; Heap, A. *A Review of Spatial Interpolation Methods for Environmental Scientists*; Geo Cat 68229; Geoscience Australia: Canberra, Australia, 2008; pp. 137–145.
11. Chilès, J.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2012.

12. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*; Springer: Berlin, Germany, 2013; Volume 103.
13. Gasmelseid, T.M.; Gasmelseid, T.M. *Handbook of Research on Hydroinformatics: Technologies, Theories and Applications (1 Volume)*, 1st ed.; IGI Global: Pennsylvania, PA, USA, 2010.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).