A new approach to targeting drilling locations: Quantifying geological knowledge in drilling campaigns using model-based design of experiments approaches

P. Deussen and F. Galvanin

University College London, United Kingdom

The performance of Kriging models depends on their semivariogram and the estimated model parameters. To maximise sampling efficiency, common sampling objectives target high expected grades, minimise Kriging variance or both. An alternative method, model-based design of experiments (MBDoE), selects sampling locations based on the sensitivity of a model to input changes, such as parameter values or locations, at which a new sample has an outsize influence on the model predictions. This maximises the information gain with respect to the model. In this paper, the efficiency of MBDoE is compared to a random design and a pitfall of MBDoE, lack of exploration, is addressed. A 2-D concentration profile was generated in-silico with a spherical correlation structure. With 20 samples, the MBDoE procedure correctly identifies the model from two candidates with 85% confidence and estimates two of three parameters correctly and with statistical significance. The random comparison fails to identify the correct model, successfully estimates only one parameter and underperforms in a variance-based exploration metric by up to 17%.

INTRODUCTION

Following over a decade of underinvestment in exploration and declining mine productivity (Canart *et al.*, 2020), the mining industry has in recent years seen higher allocations of capital for exploration, especially of precious and battery metals. These metals have once again become economic priorities, as inflation, reduced geopolitical security and industrial policy shifts make sustained higher prices of these metals more likely. One major bottleneck in the exploration for value-dense metals, such as silver or copper, is the identification of new deposits and accurate resource and reserve estimation. In this process, sampling the ground by diamond drill sampling is one of the most expensive steps. Thus, efficiently sampling these locations and inter- and extrapolating the sampled information well is crucial.

Kriging models are used in industry for this purpose. The sampling locations can be targeted using qualitative geological observations in the field and the ensuing expert interpretation by a geologist, or by varyingly dense grid patterns. Methods such as conditional simulation of Kriging models have been used to target locations that minimise the model prediction variance (Rossi and Deutsch, 2014). Recent studies have shown that multi-objective Bayesian optimisation for new drilling targets based on Kriging models can provide superior results to classical designs in a phosphate deposit (Jafrasteh and Suarez, 2021) and that model-based design of experiments (MDBoE) techniques can discriminate between possible models and improve model parameter estimates in an in-silico case study (Deussen and Galvanin, 2023). Some remaining issues include the lack of an explorative component in the objective function of the experimental design and the need for clear acceptance and rejection criteria for candidate models. Therefore, a multi-objective MDBoE (Galvanin *et al.*, 2016) approach for model discrimination, parameter estimation and reduction of prediction variance is proposed and tested in this paper.

PROPOSED PROCEDURE AND OBJECTIVES

Industry Standard in Experimental Design for Geological Exploration

In geological exploration a traditional experiment design technique is to sample in areas where the Kriging model has the highest predicted variance. In all but name, this is an MDBoE procedure as well. In the process and chemical industries, this design is known as V- or G-optimality, which aims to target locations in the design space (i.e., process operating conditions) that reduce the model variance (Fedorov, 2010; Kiefer and Wolfowitz, 1959 and 1960; Shahmohammadi and McAuley, 2019). G-optimal designs target the minimisation of the maximum prediction variance, while V-optimal designs target the minimisation of the average prediction variance. Their application is not limited to Kriging models, and it may be calculated in any parametrised model from the Fisher Information Matrix (Fisher, 1935), which will be explained in more detail in the parameter estimation section.

Another important aspect of experiment design in the mining industry is the scheduling problem. Usually, there will not be an unlimited number of drilling rigs available and the turnaround time of the assays in laboratories will also be a factor. Therefore, several sampling locations will need to be decided at once, without directly obtaining the updated information after each sample is taken. While this scheduling problem is outside the scope of this investigation, it should be investigated at a later point, as this may significantly change the optimal sampling strategy.

Objectives

The challenge with Kriging models is that areas with the highest predicted variance are not always the areas where the model has the least predictive power – it is the area where it is assumed that the model has the least predictive power. However, incorrectly selected model parameters or an altogether sub-optimal choice of semivariogram or Kriging model may result in areas of high predicted variance that are not the areas where there is the most model mismatch. Therefore, the objective of this paper is to propose a quantitative framework for the selection of the optimal model, estimation of its parameters and exploration of the design space. The geological information that the models describe will be quantified and used to select optimal samples that should improve these three objectives as the sampling (drilling out) of the prospective area advances.

Suggested Experimental Design Procedure

One iteration of the suggested experimental design procedure can be seen in Figure 1. It begins from the 'Experiment (Drilling)' box, with some initial samples that are needed to fit a preliminary model.

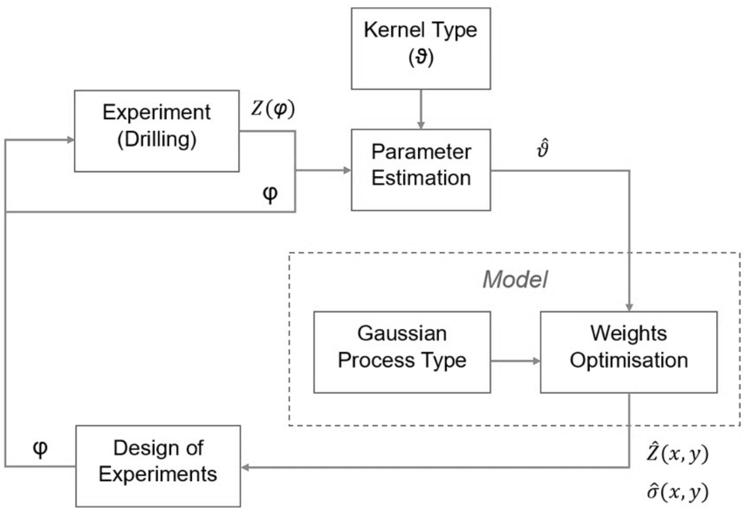


Figure 1. Flowchart of a joint experiment design method as applied to Kriging models.

The semivariogram parameters, ϑ, for each kernel (spherical and Gaussian) are estimated from existing samples in the 'parameter estimation' step. Using these estimates, for each kernel, an ordinary Kriging model (equations 1-8) is fitted with the optimal weights found based on the parameter estimates, and the correction for negative weights by Deutsch (1996). This model gives the estimates of concentration mean and variance at each location ($\hat{Z}(x, y)$ and $\hat{\sigma}(x, y)$). The multi-objective MBDoE algorithm is then used to determine at most three new samples in each 'Design of Experiments' iteration. These locations are sampled in the 'Experiment (Drilling)' stage, and the true concentrations at those points, $Z(\varphi)$, are used to update the parameter estimates. The loop is repeated until the design objectives are achieved or the sampling budget is exhausted. The constraints on the sampling budget can result from a drilling budget, seasonal or weather factors and lab turnaround times for assays that dictate the time window (number of design iterations) in which drilling will be possible.

METHODOLOGY

The methodology consists of three sections, aligned with Figure 1. The first section covers Kriging models and how they are optimally fit to sampled data. The second section covers the experimental design procedures and their application to Kriging models in a geological exploration context. This allows for the optimal targeting of new sampling locations and thus the acquisition of data points to be used in the Kriging model, outlined in section one. The third section covers criteria for measuring the performance of model predictions relative to each other and the newly acquired sampling points. It is these criteria that will be used to make final decisions on rejection or acceptance of candidate models.

Fitting Kriging Models

Equations for Optimal Fit of Candidate Kriging Models

Kriging (Sahimi, 2011) is a probabilistic technique used to create continuous spatial models given a few sample locations. They are based on random normal variables that are transformed by a correlation function, also known as a kernel. It determines the spatial correlation structure, i.e., how the probability of one location being influenced by the value of the variable at another location varies with distance and direction. A Kriging model returns an estimated mean and variance value of the variable at each unsampled location. The correlation function, $R(h)$, relates the semivariance, $\gamma(h)$, of the likelihood of variable values at unsampled locations to the distance h between them and the overall variance, σ_z^2.

$$\gamma(h) = \sigma_Z^2 - R(h) \qquad [1]$$

A commonly used variant of Kriging is ordinary Kriging. It is the best linear unbiased estimator of the mean expected concentration of the variable in question, Z, and its local variance, σ_{OK}^2. It is found using the expected square error to the mean from the estimator, $\hat{Z}$. To ensure unbiasedness, there is no need for a population variance estimate, (equations 5 & 6). The following equations describe the method.

$$\sigma_{OK}^2 = E[(Z - \hat{Z})^2] \qquad [2]$$

The estimator, $\hat{Z}$, is based on the collected samples, Z_i and the relative importance weights of these samples, w_i. Both i and j are subscripts used to denote elements of the set of sampled locations.

$$\hat{Z} = \sum_{i=1}^{N} w_i Z_i \qquad [3]$$

The optimal values of the weights w_i and w_j can be found by substituting equation 3 into equation 2:

$$\sigma_{OK}^2 = \sigma_Z^2 - 2\sum_{i=1}^{N} w_i R(Z, Z_i) + \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j R(Z_i, Z_j) \qquad [4]$$

In ordinary Kriging, to avoid estimating the variance of the total population from the subset of sampled locations, the following constraint ensures that the weights sum to one and thus include all possible information and do not rely on a population variance estimate from previous samples.

$$\sum_{i=1}^{N} w_i = 1 \qquad [5]$$

Substituting equations 1 and 5 (the constraint) into equation 4 results in one degree of freedom:

$$\sigma_{OK}^2 = 2\sum_{i=1}^{N} w_i\, \gamma(Z, Z_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\, w_j\, \gamma\left(Z_i, Z_j\right) \quad [6]$$

Then, $\partial\, \sigma_{OK}^2\ /\ \partial\, w_i = 0$, i.e., an optimisation of the fit is carried out that determines the values of the weights which minimise the Kriging variance, giving zero degrees of freedom. For all locations *j*:

$$\sum_{i=1}^{N} R(Z_i, Z_j)\, w_i = R(Z, Z_j) \quad [7]$$

From equation [7] a linear system of equations which can be written as a matrix of weights, **W**, with the dimensions of sampled locations, n_s, by unsampled locations, n_u. The matrix **P** (also n_s x n_u -dimensional) holds the correlations between the sampled and unsampled locations, and the matrix **A** (n_s x n_s -dimensional) holds the correlations between sampled locations in the rows and the columns.

$$\mathbf{W} = \mathbf{A}^{-1}\mathbf{P} \quad [8]$$

Then, the weights in matrix **W** can be substituted into equations 3 and 6 to obtain a mean and variance estimate of the random variable at all unsampled locations. Nothing in the above calculations prevents weights from taking negative values, for instance when one data point is shielded by another closely behind it. These often yield nonphysical predictions of the variable and create numerical difficulties in computational programmes, so they are corrected using the algorithm suggested by Deutsch (1996).

Choice of Semivariogram
Since the semivariogram, or kernel, describes the correlation structure between the random variable at different unsampled points, it is a crucial component in building a Kriging model. Two semivariograms which are common in industry and have structurally identifiable parameters are used in this study.

- Gaussian (Rossi and Deutsch, 2014)

if $h < r$

$$\gamma(h) = (s)\left(1 - e^{-3\frac{h^2}{r^2}}\right) + n \quad [9]$$

if $h > r$

$$\gamma(h) = s + n \quad [10]$$

where $\gamma(h)$ is the semivariance as a function of distance h, n the nugget effect, s the sill and r the range. After the range is exceeded, the semivariance is constant, i.e., two points further apart do not influence each other's likely concentrations other than belonging to one population with the same variance.

- Spherical (Sahimi, 2011)

if $h < r$

$$\gamma(h) = (s)\left(\frac{3h}{2r} - \frac{1}{2}\left(\frac{h}{r}\right)^3\right) + n \quad [11]$$

if $h > r$

$$\gamma(h) = s + n \quad [12]$$

These two kernels can be substituted into equation 1 to construct Kriging models from them. In this case study, the in-silico data is generated with an isotropous correlation structure. This makes the lag angle irrelevant. The tolerance is also specified with respect to the in-silico data set in the case study section.

Design-of-Experiments Procedures
Design of Experiments (DoE) is a statistical methodology used to systematically plan, conduct, and

analyse experiments to optimise the assessment of a process or system. Model-based techniques (Franceschini and Macchietto, 2008), unlike space-filling or black-box designs, require a model structure and knowledge of its parameters to vary them and assess their impact.

Model-Based Design of Experiments for Parameter Precision

Model-based designs can have several objectives, including the precise estimation of model parameters. The expected information gain of a sample at a coordinate with respect to the estimability of the model parameters is quantified and the optimal samples chosen. This quantification is based on the local sensitivities (partial derivatives) of model predictions, $\hat{\mathbf{y}}_i$ (1, …, n_m), to variations in the model parameters, $\boldsymbol{\vartheta}$ (1, …, n_ϑ). They are recorded in a sensitivity matrix **Q** (Franceschini and Macchietto, 2008). In a geological context, the model outputs can include rock properties or concentrations while the model parameters in Kriging models consist of the semivariogram parameters. At all locations on the grid, except at the sampled points (1, …, n_{exp}) where the model sensitivities are zero, there exists:

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial \hat{y}_1}{\partial \vartheta_1} & \cdots & \frac{\partial \hat{y}_1}{\partial \vartheta_{n_\vartheta}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{y}_{n_m}}{\partial \vartheta_1} & \cdots & \frac{\partial \hat{y}_{n_m}}{\partial \vartheta_{n_\vartheta}} \end{bmatrix} \approx \begin{bmatrix} \left(\frac{y_1 - \hat{y}_1}{\vartheta_1 - \hat{\vartheta}_1}\right) & \cdots & \left(\frac{y_1 - \hat{y}_1}{\vartheta_{n_\vartheta} - \hat{\vartheta}_{n_\vartheta}}\right) \\ \vdots & \ddots & \vdots \\ \left(\frac{y_{n_m} - \hat{y}_{n_m}}{\vartheta_1 - \hat{\vartheta}_1}\right) & \cdots & \left(\frac{y_{n_m} - \hat{y}_{n_m}}{\vartheta_{n_\vartheta} - \hat{\vartheta}_{n_\vartheta}}\right) \end{bmatrix} \quad [13]$$

The parameter is perturbed by 5% to approximate the derivative. The matrix **Q** can be adjusted by the uncertainty from the Kriging variance at each point (Deussen and Galvanin, 2023), resulting in more precise estimates but also the danger of local optima. Hence it will not be used in this procedure. **Q** is combined with the sampling error, σ, to find the covariance and correlation matrices between model parameters, $\mathbf{V}^{\boldsymbol{\vartheta}}$ and **C**, respectively. The element $k,l \in n_\vartheta$ and shows the correlation between ϑ_k and ϑ_l.

$$\mathbf{V}^{\boldsymbol{\vartheta}} = [\mathbf{Q}^{\mathrm{T}} \sigma^{-2} \mathbf{Q}\,]^{-1} \quad [14]$$

$$C_{kl} = \frac{V^{\vartheta}{}_{kl}}{\sqrt{V^{\vartheta}{}_{kk} V^{\vartheta}{}_{ll}}} \quad [15]$$

The elements (correlation values), C_{kl}, range from -1 (total anticorrelation) over 0 (no correlation) to 1 (total correlation). Critical values exceed 0.95 and -0.95, where parameter estimation is difficult or impossible. The most convenient and comprehensive way to store all of this information together is the Fisher Information Matrix, **H** (Fisher, 1935). It is the best basis for experiment design criteria.

$$\mathbf{H} = \sum_{i=1}^{n_{exp}} \sum_{j=1}^{n_m} \left[\frac{1}{\sigma_{ij}^2} Q_{ij}{}^{\mathrm{T}} Q_{ij}\right] \quad [16]$$

The objective must be a scalar measure of **H**, which meaningfully includes the information across all locations and modelled attributes. Three optimality criteria exist: A-, D- and E-optimality (in addition to the less commonly used G- and V-optimality and others for specialised objectives). A-optimal designs minimise the trace of the inverse of **H**. D-optimal designs maximise the determinant of **H** and E-optimal designs maximise the smallest eigenvalue of **H** (Franceschini and Macchietto, 2008).

Model-Based Design of Experiments for Model Discrimination

For the case that several models appear likely, model discrimination criteria of varying complexity exist. The most fundamental model discrimination criterion, D_{HR}, was suggested by Hunter and Reiner (1965) and is conceptually simple to understand. It is found from the difference between model predictions, $\hat{Z}_m$ (at a coordinates x,y and a set of parametric inputs ϑ) only. For multiple models m and n:

$$D_{HR}(x, y, \vartheta) = \sum_{m=1}^{M-1} \sum_{n=m+1}^{M} \left[\hat{Z}_m(x, y, \vartheta) - \hat{Z}_n(x, y, \vartheta)\right]^2 \quad [17]$$

This was extended by Buzzi-Ferraris *et al.,* (1984, 1990) to include the model variance, such that areas with a high variance had lower predictive power in the discrimination between models. Schwaab *et al.*, (2006) further extended this to where models with low probability of being true are penalised in determining the optimal sampling location for model discrimination. The original authors use the chi-squared probability, which effectively measures the goodness of fit of model predictions at each point when repeating the experiment several times. This is well-suited only to frequentist scenarios. The case of sampling the ground, however, is Bayesian, as once a location has been sampled there is no longer any uncertainty about that location in the frequentist sense (i.e., repeatedly sampling the same location will reveal minimal new information about the Kriging model). Therefore, the probability term is modified (equations 18-20) to use the inverse total Kriging variance of any model m for all species k, parameters σ and locations (x,y). The larger the total variance, the more uncertain the predictions.

$$D_{MS}(x,y,\vartheta) = \sum_{m=1}^{M-1}\sum_{n=m+1}^{M}(P_m,P_n)\left[\hat{\mathbf{Z}}_m(x,y,\vartheta) - \hat{\mathbf{Z}}_n(x,y,\vartheta)\right]^{\mathrm{T}}\mathbf{V}_{m,n}^{-1}(x,y,\vartheta)\left[\hat{\mathbf{Z}}_m(x,y,\vartheta) - \hat{\mathbf{Z}}_n(x,y,\vartheta)\right] \quad [18]$$

$$\phi_m(\vartheta) = 1 \Big/ \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{k=1}^{K} V_m(x,y,k,\vartheta) \quad [19]$$

$$P_m(\vartheta) = \frac{\phi_m(\vartheta)}{\sum_{m=1}^{M}\phi_m(\vartheta)} \quad [20]$$

This measure of probability penalises both models that are overfit, resulting in high variance spikes between the sampled locations, as well as models that generally have a high uncertainty, with fewer spikes but a higher average uncertainty.

Multi-objective Optimisation for Joint Experiment Designs

To build an algorithm that jointly maximises three information criteria, the problem is written as a constrained optimisation based on the joint design with two objectives in Galvanin *et al.*, (2016).

$$\phi^{ID} = arg\max\{\Psi^{MD}(\varphi)\}_{\phi\in D} \quad [21]$$

s.t.

$$\Psi^{PE} = \sum_{j=1}^{N_M}\left\|\mathbf{H}_j(\varphi)\right\|/N_M \leq \varepsilon_1 \quad [22]$$

s.t.

$$\Psi^{EXP} = \sum_{j=1}^{N_M}\hat{\sigma}_j(\varphi)/N_M \leq \varepsilon_2 \quad [23]$$

where:

$$\varepsilon_1 = \max(\Psi^{PE}) \quad s.t. \quad \max(\Psi^{MD}) \quad [24]$$

$$\varepsilon_2 = \max(\Psi^{EXP}) \quad s.t. \quad \max(\Psi^{MD}), \max(\Psi^{PE}) \quad [25]$$

Where ϕ^{ID} is the overall identification criterion that includes the three objectives for model discrimination, Ψ^{MD}, parameter estimation, Ψ^{PE}, and exploration, Ψ^{EXP}. **H** is the FIM (equation 16) and $\hat{\sigma}_j$ is the Kriging variance at a location, averaged between candidate models. The solution to this problem is found using the improved epsilon constraint method (Mavrotas, 2009), which ensures that none of the suggested solutions on the pareto front are weakly efficient (equations 24 and 25). It ensures that the trade-off between the three objectives leaves no room to improve one objective without losing optimality on another. A weakly efficient solution is one that maximise one objective but leaves slack in another objective, which could be improved at no cost to the other function, which is to be avoided.

Model Performance Criteria

Kullback-Leibler Divergence

The concept of entropy is often used in machine learning applications to compare the information content of population distributions. A type of information measure to discriminate between two distributions is the Kullback-Leiber divergence (Kullback and Leibler, 1951), which can be written as:

$$D_{KL}(P_2|P_1) = \log\left(\frac{\sigma_1}{\sigma_2}\right) + \frac{{\sigma_2}^2+(\mu_2-\mu_1)^2}{2{\sigma_1}^2} - \frac{1}{2} \quad [26]$$

Where μ and σ represent the mean and standard deviation of two probability distributions. P_1 represents the prior (before sampling) and P_2 represents the posterior distribution (after sampling). The metric measures the information gain (entropy decrease) from these samples with respect to the models and is thus asymmetric. This divergence can be found for every location where there is no sample (i.e., only a model prediction). Newly sampled locations result in an infinitely large Kullback-Leibler divergence, as the entropy decrease from a distribution to a point is infinitely large, i.e., the information gain from a distribution to a point is infinite. Sampled points no longer hold any information about the models either, as there is, intuitively, no information gain in a point already known with certainty. The average of all coordinates is taken for each model and each design iteration, and is then aggregated in a moving average over the iterations. If this falls below a threshold such as 90%, a model can be rejected, as it is responsible for over 90% of the adjustments in the divergence between the distributions. The design procedure is complete once one model is left or upon exhaustion of the sampling budget.

Student t-Test
The Student t-test is a metric of the precision of parameter estimates (Franceschini and Macchietto, 2008), i.e., describing its statistical significance. The t-value of an estimate, t_i, must be compared to a reference value for the chosen confidence interval and the degrees of freedom, d_f, that are present in an estimation.

$$t_i = \frac{\hat{\vartheta}_i}{\sqrt{v_i^{\vartheta}}} \quad [27]$$

$$d_f = n_s - n_\vartheta \quad [28]$$

A precise estimate is not necessarily accurate, i.e., they may diverge from the true value. As this is an in-silico study, the estimates will also be compared to the ground truth to check for accuracy.

CASE STUDY

Definition of Ground Truth and Candidate Models
An in-silico concentration profile of a fictional, isotropic metal deposit is created using the spherical kernel (equation 11). The ground truth concentrations are generated using the parameters in Table I. Two candidate models are suggested to describe this concentration. All candidate models are ordinary Kriging, however model 1 (M1) uses a spherical kernel (equations 11 & 12) while model 2 (M2) uses a Gaussian kernel (equations 9 and 10).

Table I. Parameters used to generate ground truth values in M1 with spherical kernel (equation 11)

Parameter (ϑ)	Value
range (r)	17 hm
sill (s)	7 $(g/t)^2$
nugget effect (n)	1 $(g/t)^2$
Area	5.1 km x 5.1 km

Definition of Sampling Procedure and Objective Functions
The design variables, stored in the design vector φ, are the x- and y-coordinates of the new sample points. The kernel parameters ϑ are to be determined with the lag distance fixed to 3.8 km. The procedure will begin from five randomly selected samples. New samples will be selected according to the proposed procedure (Figure 1). At most three samples will be selected per design iteration, of which one will be located at each extreme of the pareto front, i.e., optimise one of the three objectives (equations 21-25). If the suggested locations are too close together and the grid resolution cannot distinguish them,

fewer than three samples may be added per iteration. The three objectives are the modified Schwaab *et al.*, (2006) criterion for model discrimination (equations 18-20), the D-optimal discrimination criterion for parameter estimation, and reduction of the maximum Kriging variance for exploration. The sampling budget allows for five design iterations, resulting in fifteen additional or twenty total samples.

The sampling technique used as a comparison was random sampling, since if an exploration team of geologists holds a certain set of assumptions that turn out to be false, mathematically this sampling can be represented well as being random. It may even have a bias towards clustering in sub-optimal locations. Conversely, correct assumptions about the geology would have a bias to cluster in very prospective or informative regions, so random sampling is neutral to underlying bias.

The semivariogram fitting was carried out using the Python function 'Optimise.Curvefit' from the library Scipy, with initial guesses of r = 25 hm, s = 10 $(g/t)^2$ and n = 2 $(g/t)^2$. The remaining calculations were carried out in a program in Python that was written for this purpose.

RESULTS

The goal of this objective was to drill for information that could distinguish the two models, M1 (spherical kernel) and M2 (Gaussian kernel). The ground truth and new samples are shown in Figure 2a, while the population distributions of the fitted models are shown in Figure 2b. The population distribution of M1 after 20 samples with the MBDoE method resembles the ground truth most closely.

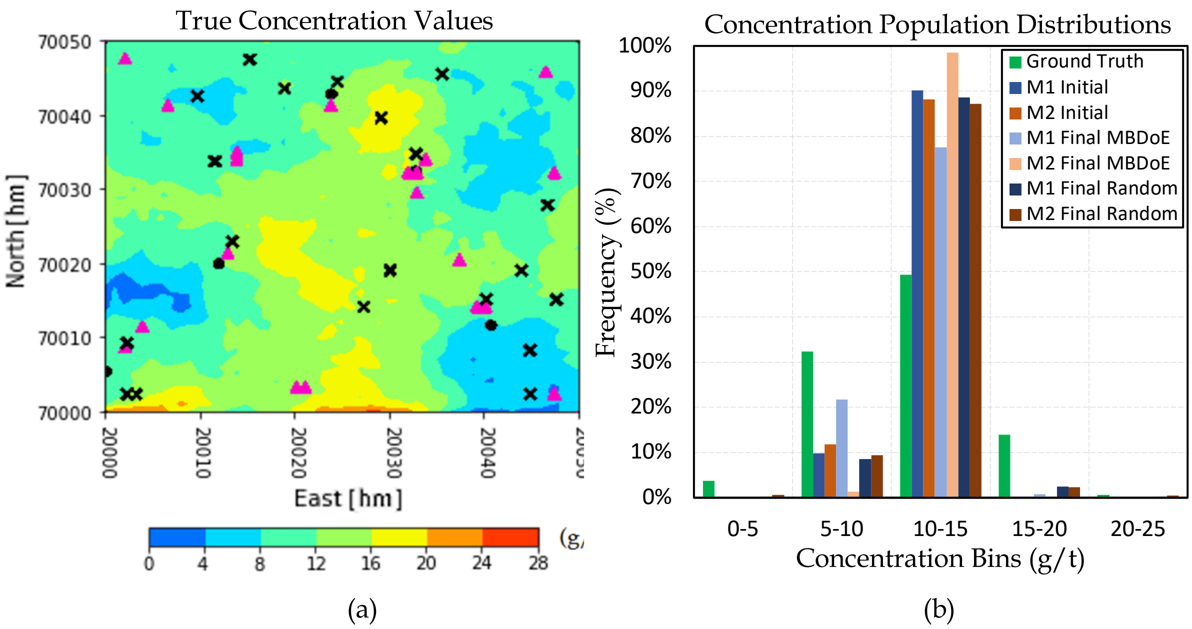


Figure 2. Map of 2-dimensional deposit generated using M1 with five random initial samples (black dots), fifteen additional samples from MBDoE (purple triangles) and fifteen alternative random samples (black crosses) (a); and population distributions of the ground truth and candidate models throughout the design procedure (b). 'Initial' refers to the state of each model with only the five initial samples while 'final' refers to the state of the model with 20 samples, obtained either by MBDoE or random sampling as indicated.

Model Identification

The MBDoE approach returns M1 as the correct model to accept within an 85% confidence interval on the third design iteration (see Figure 3), however, the moving average falls just short of the 90% confidence interval (on the third and fifth design iterations). The random sampling in comparison achieves no significant or consistent discriminating power through all five iterations, and predicts the incorrect model to be more likely at the fifth iteration. It can thus be said that the MBDoE procedure achieves its goal of outperforming the random sampling in identifying the correct model.

The Kullback-Leibler divergence (equation 26) is an accurate reflection of the model's predictive capability, as a higher change in this divergence indicates that a design iteration added new information not previously included in the model. It is significantly more sensitive and a better objective quantification of the geological information in a model than analysing the change in model mean and variance separately, as has been done in past work (Deussen and Galvanin, 2023). One drawback is that it loses the information at locations that were sampled in each iteration. This metric could be complemented by the chi-squared statistics of these locations to re-incorporate their information.

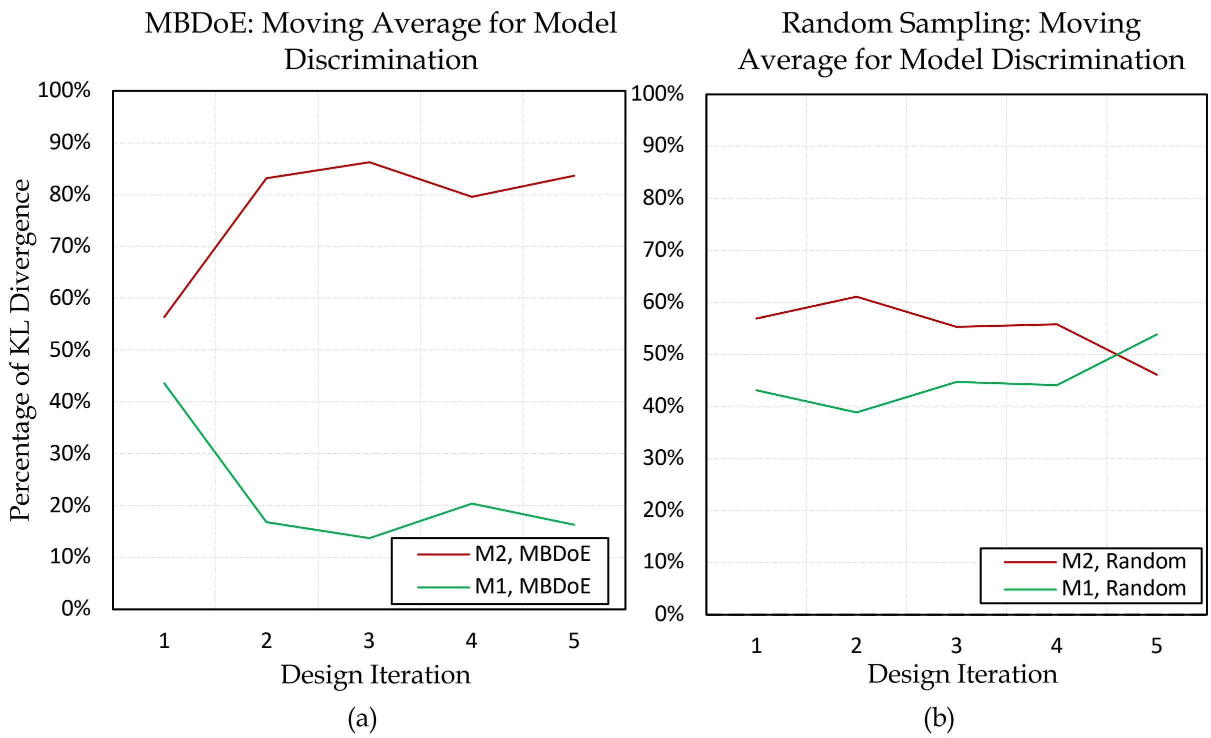
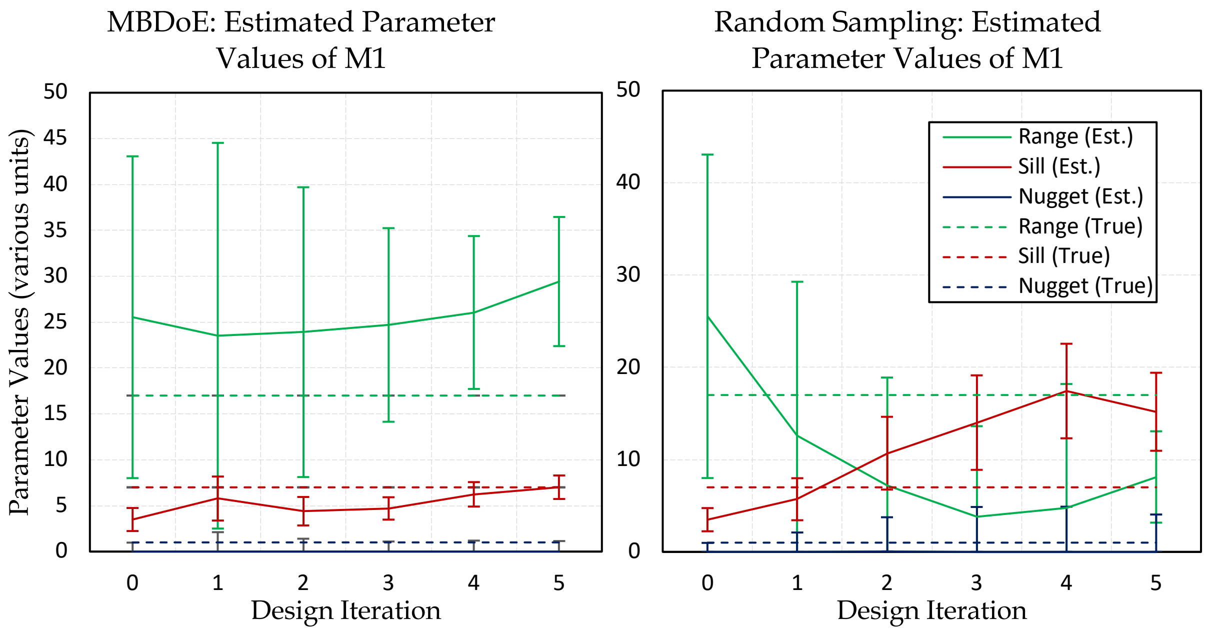


Figure 3. Model discrimination metric: moving averages of model contributions to Kullback-Leibler divergence for MBDoE (a) and random sampling (b).

In summary, even though it accepts M1 within only an 85% confidence interval, the MBDoE procedure coupled with the KL-divergence as a metric succeed at enhancing the discriminatory power of experimental designs and providing an objective metric to track its progress.

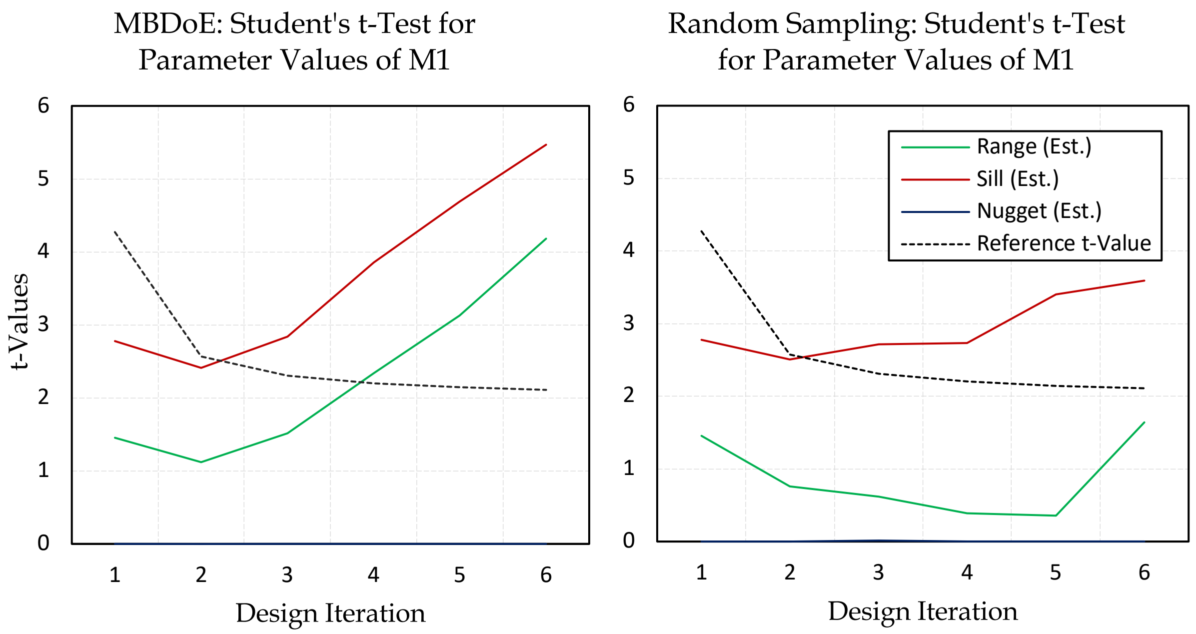
Parameter Estimation

The parameter estimation carried out in the MBDoE approach (see Figure 4) correctly returns two parameters (sill and nugget) within their margins of error. The Student t-Test (see Figure 5) is passed for both the range and the sill, even though the range parameter is incorrectly estimated. The random sampling comparison returns only the nugget correctly within a margin of error almost four times higher than the error for the same parameter in the MBDoE method. The Student t-Test is passed for only the sill value, although the parameter estimate does not reflect the ground truth value within the margin of error. The improvement in t-values is steeper in the MBDoE procedure.



(a) (b)

Figure 4. Parameter estimates and true values for MBDoE (a) and random sampling (b).



(a) (b)

Figure 5. Student t-Test values for parameter estimates for MBDoE (a) and random sampling (b); note that the nugget effect has t-values close to zero and is thus not easily visible on the graph.

The low t-values of the nugget result from the low estimate value, which penalises error margins harshly in terms of statistical significance. Either more design iterations or a dataset generated using a larger nugget effect are needed to further examine its estimability. The MBDoE procedure sinks into a local optimum with the range estimate. The error decreases while the estimate diverges from the ground truth. In summary, the sampling budget is insufficient to estimate all parameters in either procedure. However, the MBDoE procedure is a useful tool to improve the statistical significance of estimates in only a few design iterations and in obtaining values closer to the ground truth than random sampling. In addition, the MBDoE method combined with the Student t-test can quantify the uncertainty of geological parameters and can guide exploration decisions.

Exploratory Sampling

The objective of the exploratory component is to avoid local optima for the other two objectives and to reduce the prediction variance. Only one parameter diverges from the ground truth in the MBDoE procedure in five design iterations, which is an improvement over past work (Deussen and Galvanin, 2023) on divergence and the number of samples necessary to reach these estimates. The MBDoE procedure is also more effective at avoiding local optima for parameter estimation than the random procedure (compare Figures 4a and 4b). As an independent metric of the exploration component (see Figure 6), the average variance as a percentage of the sill parameter at any design iteration can be used, as the sill, as this dictates the maximum Kriging variance in a model.

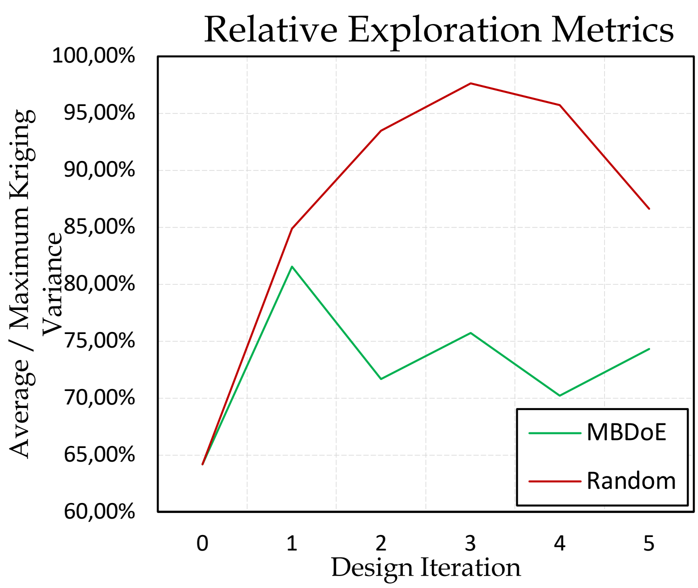


Figure 6. Average Kriging variance as a percentage of maximum Kriging variance as exploration metric.

The variance reduction component is thus also effective at achieving its goal in the MBDoE procedure vs the random procedure, achieving a result that is 12-25% better, depending on the design iteration. Finally, even though clustering is a feature of MBDoE applied to Kriging models (see Figure 2) and there are still local optima in the parameter estimation, the exploration component fulfils its stated objective.

CONCLUSIONS

A multi-objective MBDoE approach was compared to a random sampling approach of an in-silico prospect. Out of two candidate models, the MBDoE procedure successfully identified the ground truth model upon exhaustion of the drilling budget. The MBDoE discrimination objective performed ca. 40% better than the random sampling and the Kullback-Leibler divergence proved a sensitive and robust discrimination metric. The parameter estimation is the least successful of the three objectives, as not all parameters are estimated to statistical significance and convergence with ground truth values. However, the MBDoE procedure outperforms random sampling in both accuracy and precision. The exploration component reduced the instances of local optima in estimation and discrimination functions, as well as reducing the variance more than random sampling. Overall, the MBDoE approach showed superior results in all objectives and succeeded at quantifying the information in robust metrics. The precision of model discrimination and accuracy of parameter estimates should be refined. Monte Carlo simulations can be used to better estimate parametric uncertainty.

REFERENCES

Buzzi-Ferraris, G., Forzatti, P., Emig, G. and Hofmann, H. (1984) Sequential experimental design for model discrimination in the case of multiple responses. *Chem. Eng. Sci.*, 45 (2), pp 81 – 85.

Buzzi-Ferraris, G., Forzatti, P. and Canu, P. (1990) An improved version of a sequential design criterion for discriminating among rival multiresponse models. *Chem. Eng. Sci.*, 45 (2), pp 477 – 481.

Canart, G., Kowalik, L., Moyo, M. and Ray R.K. (2020) Has global mining productivity reversed course? *McKinsey & Company MineLens Commentary*.

Deussen, P. and Galvanin, F. (2023) A joint model-based design of experiments approach for the identification of Kriging models in geological exploration. *Proceedings of the 33rd European Symposium on Computer Aided Process Engineering*, Antonis Kokossis, Michael C. Georgiadis, Efstratios N. Pistikopoulos (eds.). European Federation of Chemical Engineering, Athens, Greece.

Deutsch, C.V. (1996) Correcting for negative weights in ordinary Kriging. *Computers and Geosciences*, 22(7), pp 765-773.

Fedorov, V. (2010) Optimal experimental design. *WIREs Comp. Stat.*, 2, pp 581-589.

Fisher, R.A. (1935) *The Design of Experiments*. Oliver & Boyd, Edinburgh.

Franceschini, G. and Macchietto, S. (2008) Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63, pp 4846-4872.

Galvanin, F., Enhong, C., Al-Rifai, N., Gavriilidis, A. and Dua, V. (2016) A joint model-based design approach for the identification of kinetic models in continuous flow laboratory systems. *Computers and Chemical Engineering*, 95, pp 202-215.

Hunter, W.G. and Reiner, A.M. (1965) Designs for discriminating between two rival models. *Technometrics*, 7 (3), pp 307-323.

Jafrasteh, B. and Suarez, A. (2021) Objective functions from Bayesian optimization to locate additional drillholes. *Computers and Geosciences*, p 147.

Kiefer, J. and Wolfowitz, J. (1959) Optimum designs in regression problems. *Ann. Math. Stat.*, 30(2), pp 271-294.

Kiefer, J. and Wolfowitz, J. (1960) The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12, pp 363-366.

Kullback, S. and Leibler, R.A. (1951) On information and sufficiency. *Ann. Math. Statist*, 22(1), pp 79-86.

Mavrotas, G. (2009) Effective implementation of the ε-constraint method in multi-objective mathematical programming problems. *Appl. Math. Comp.*, 213, pp 455-465.

Rossi, M.E. and Deutsch, C.V. (2014) *Mineral Resource Estimation*, 1st etdn. Springer, Heidelberg.

Sahimi, M. (2011) *Flow and Transport in Porous Media*. 2nd etdn. Wiley-VCH Verlag, Weinheim.

Shahmohammadi, A. & McAuley, K. (2019) Sequential model-based A- and V-optimal design of experiments for building fundamental models of pharmaceutical production processes. *Computers and Chemical Engineering*, 129.

Schwaab, M., Silva, F.M., Queipo, C.A., Barreto, Jr., A.G., Pinto, J.C. (2006) A new approach for sequential experimental design for model discrimination. *Chem. Eng. Sci.*, 61, pp 5791-5806.

Philipp Adrian Deussen

PhD Candidate
University College London (UCL)

Philipp has a master's degree in chemical engineering, has worked in a mergers and acquisitions company, with an economics think tank and taken geology and finance courses. His initial research work focussed on optimal parameter estimation in reaction kinetics using model-based design of experiments methods. The focus of his PhD is to apply these methods to optimise exploratory drilling in the mining industry by better identifying model parameters and discriminating between models of the subsurface. His research interests also include applying and developing machine learning models to geological systems.