AN INFORMATION-THEORETIC SAMPLING STRATEGY FOR THE RECOVERY OF GEOLOGICAL IMAGES: MODELING, ANALYSIS, AND IMPLEMENTATION

Thesis Exam: Doctorado en Ingeniería Eléctrica

Felipe Santibáñez-Leal

Facultad de Ciencias Físicas y Matemáticas,
Universidad de Chile

IDS Group

RoadMap

Context

Regionalized Variable

Set of random variables with spatial dependence and spatial correlation.

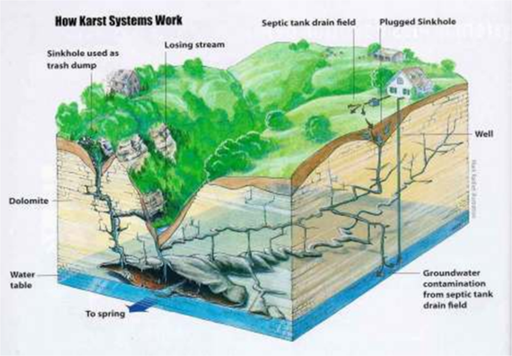


Figure 1: Example of $3D$ Regionalized variable

Figure 2: Example of $2D$ Regionalized variable

Context

Sampling

Only a small amount of measurements (Well) are available. Each Well is Extremely Expensive.

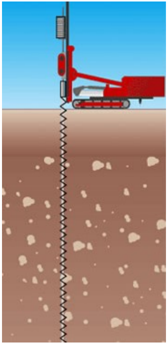


Figure 3: Sampling Scheme

Figure 4: Actual Well

Context

Inverse Problem

With available measurements an inference system is required.

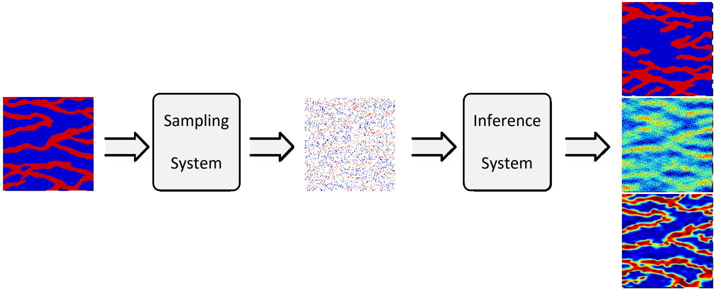
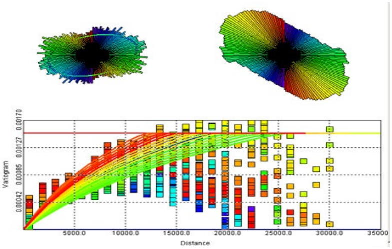


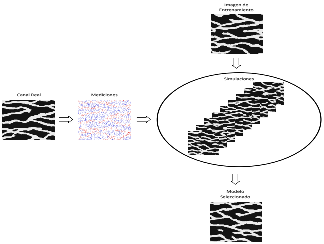
Figure 5: General Inverse Problem Scheme in Regionalized variables characterization

Conventional methods

Acquisition		Inference Scheme
• High	→	Estimation (variograms, covariograms)
• Low	→	Simulations (MPS)



(a) Variographic analysis

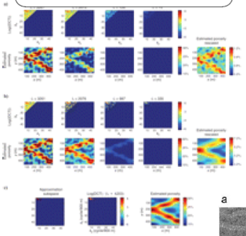


(b) MPS scheme

Figure 6: Classical Approaches

New approaches in Geostatistics

Jafarpour et al. 2009



Calderon et al. 2016

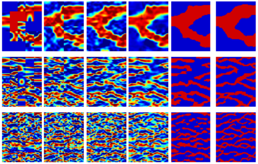


Figure 11: Reconstructions as a function of the block-size. First row presents the reconstruction of single-channel structure (0.1% sampling ratio) and second row a multichannel model (0.7% sampling ratio). From left to right by column: 40 x 40pp, 100 x 100pp, full image, average recovered image and average reconstruction after thresholding.

Calderon et al. 2014

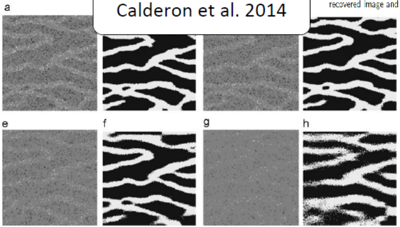


Fig. 12. Average reconstruction across block sizes with different proportion of measurements. (a) and (b) show the 10% of the data and the reconstruction, respectively. (c) and (d) show the same graphics for 8% of data. (e) and (f) for the 5%. (g) and (h) for the 2%.

Bias and Preferential Sampling

Non Preferential Sampling

- Uniform random sampling
- Deterministic stratified sampling
- Randomized stratified sampling
- Multiscale stratified sampling

Bias and Preferential Sampling

Non Preferential Sampling

- Uniform random sampling
- Deterministic stratified sampling
- Randomized stratified sampling
- Multiscale stratified sampling

Preferential sampling, Why not?

- Maximum indicator sampling

Main Questions

Relevant Questions

- Given K available measurements, what is the *best* location for each one?
- Given an statistical image model and a predefined number of measures, is there an optimal sampling scheme?
- What is the best inference methodology for the proposed optimal sensing strategy?
- Under the use of *MPS* approaches, can entropy and mutual information be good criteria for optimal sampling?
- What is the relation between the complexity of the model and the sampling process under the context of *MPS* approaches?

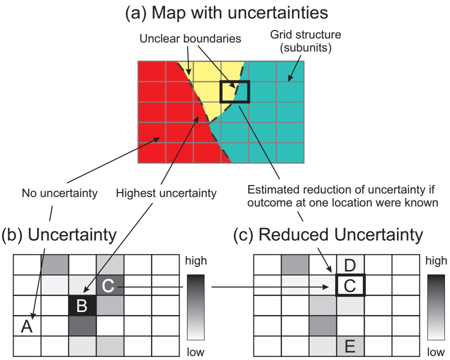
Motivation: $2D$ Regionalized Variable

Information Theory

Wellman [16] proposed the use of **Information Theory** to geostatistical analysis. They used **Conditional Entropy**.

Information in Regionalized Variables

- Each position represents a **random variable** (with spatial dependence).
- Transitions are zones of **high uncertainty**. Then, measuring these zones will reduce global uncertainty.
- Wellman uses **joint probabilities**.



Hypotheses

Main hypotheses:

- In *Geostatistics*, at low sampling regimes, the incorporation of prior information (based on *MPS* and training images) in the design of sampling strategy improves the performance with respect to classical sampling approaches.
- Adaptive sensing schemes can be integrated in the inference to improve the state-of-the-art of geological field characterization.
- Information measures are accurate predictors of the complexity of simulation tasks, and can be used to improve inference for the type of decisions carried out in planning and production stages.

Objectives

Main Objective

Enhance the reconstruction of images describing *2-D* categorical regionalized variables by the use of new *sensing* strategies that takes into account uncertainty reduction under an adaptive strategy by taking advantage of its spatial structure and other sources of expert knowledge of the media of interest.

Objectives

Specific objectives:

- Formalize a general theoretical framework for optimal sampling design with focus on, but not limited to, categorical variables.
- Develop an adaptive sensing design framework using *joint entropy* and *mutual information* to measure uncertainty and spatial structure.
- Compare the performance of the full combinatorial sampling strategy with the sequential and the adaptive sequential strategies within sampling design framework for the optimal information decision task.
- Study some stopping criterion for each sampling strategy as a function of the capacity of the field as a measure of its complexity.
- Evaluate proposed sampling strategy on Markov random field models on finite alphabet for regionalized random variables as a controlled scenario to validate the proposed method.
- Evaluate the proposed sampling strategy in a practical realistic context of grade control for short-term planning.

Formalization

Binary Regionalized Variable

- We formalize this problem considering *2-D* variables with spatial correlation
- Regionalized variables arises naturally as a suitable model to represent *2-D* random fields (finite alphabet images) describing the subsurface channels.
- A regionalized variable Z is a square *2-D* random array of variables representing a discrete image of finite size $M \times M = N$, consisting of M^2 discrete random variables:

$$Z_{u,v} : (\Omega, \mathbb{P}) \rightarrow \mathcal{A} = \{0, \ldots, |\mathcal{A}| - 1\} \quad \forall (u,v) \in \{1, \ldots, M\}^2, \tag{1}$$

with values in the finite alphabet $\mathcal{A}$.

Formalization

Binary Regionalized Variable

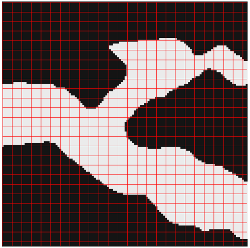
- Without loss of generality, the random field Z can be rearranged as a finite dimensional vector X in $\mathbb{R}^N$
- The object to be characterized is a random image (or random field) denoting the subsurface distribution by a collection of finite alphabet random variables $X = \{X_i : i \in [N]\}$ with, $[N] = \{1, \ldots, N\}$.
- For every position i in the array, X_i is a random variable with values in the finite alphabet $\mathcal{A}$

Formalization

Binary Regionalized Variable

- Then we can define the collection X_I as the subset of X_i variables with $i \in I$, where I represent any subset of $[N^2]$,
  $X_I = \{X_i : i \in I\}$
- In addition, the object X^I is defined as the complement of X_I over the collection X
- $X^I = \{X_i : i \in [N^2] \setminus I\}$

X_{11}	X_{12}	$\dots$	X_{1n}
X_{21}	X_{22}		$\vdots$
$\vdots$		$\ddots$	
X_{n1}	$\dots$		X_{nn}



Problem Formalization

Entropy and Uncertainty of X

- The probability density function (*pdf*) of X_i is denoted by $\mathbb{P}_{X_i}$ in $\mathcal{A}$.
- The collection X is equipped with its joint probability distribution that we denote by $\mathbb{P}_X$ in $\mathcal{A}^N$.
- As a short hand, X y $\mathbb{P}_X$ denote the vectorized random field and its joint probability, respectively.

Problem Formalization

Optimal Sensor Placement *OSP*

- Thus, the problem of *OSP* can be posted as the problem of selecting a subset of K elements of $[N]$.
- Let $\mathbf{F}_K \equiv \{f : f \subset [N], |f| = K\}$ be the collection of functions that select K-elements from N candidates.
- Every $f \in \mathbf{F}_K$ is a measurement allocation rule that models the process of measuring the positions $f(1), f(2), \ldots, f(K)$ in the random field.

Problem Formalization

OSP by uncertainty reduction

- Adopting the concept of entropy as a measure of uncertainty of a random variable [4], we propose an algorithm that finds the placement rule f through optimal reduction of *a posteriori* entropy.
- The criteria used in Eq. (2) states that the measurement of the most uncertainty set of K positions will provide an optimal global reduction of the uncertainty for the media of interest (from the point of view of *information theory*).

$$X_f^* = \arg\max_{X_f \subset X} H(X_f) \tag{2}$$

$$H(X^f|X_f) = H(X) - H(X_f) \tag{3}$$

Summary

Concepts

- Scenario: $X = \{X_i | (i) \in \{1, \ldots, N^2\}\}$
- Objective: $H(X_{\text{No Measured}} | X_{\text{Any set of Measures}}) \geq H(X_{\text{No Measured}} | X_{\text{OSP set of Measures}})$
- OSP algorithm
  - $\arg\max_{X_{\text{Measured}}} \; H(X) - H(X_{\text{No Measured}} | X_{\text{Measured}})$
  - $\arg\min_{X_{\text{Measured}}} \; H(X_{\text{No Measured}} | X_{\text{Measured}})$
  - $\arg\max_{X_{\text{Measured}}} \; H(X_{\text{Measured}})$

AMIS Proposal

Sampling Strategies for Uncertainty Reduction in Categorical Random Fields: Formulation, Mathematical Analysis and Application to Multiple-Point Simulations

AMIS Formalization

It is possible to show with some formality that the entropy $H(X)$ has an operational meaning for the task of simulating X using n i.i.d. realizations

Entropy as an Indicator of Simulation Complexity

Considering ϵ sufficiently small, for all $(x_1, .., x_n) \in B_n(\epsilon)$, then:

$$\mu_X^n(x_1, .., x_n) \approx 2^{-n \cdot \mathcal{H}(\mu_X)} \approx \frac{1}{|B_n(\epsilon)|}.$$

Then within this set $B_n(\epsilon)$, which is typical, all its elements have the same probability. This means that when making i.i.d. samples of the model μ_X^n and n is sufficiently large, a single sample of this typical set (that happens with very high probability), has the same probability than any other element of the set.

AMIS Formalization

Problem Setting

The optimal sampling problem can be posted as a minimum cost decision problem, where the cost is the complexity to characterize a random object in terms of i.i.d. samples.
More formally, for a given number $k \leq M^2$ of positions to be sensed in the pixel-domain $[M] \times [M]$, let $\mathbf{F}_k \equiv \{f : \{1, .., k\} \rightarrow [M] \times [M]\}$ be the collection of functions that select k-elements from $[M] \times [M]$. Every $f \in \mathbf{F}_k$ represents a sampling rule of size k that defines the positions to be sensed in the field, denoted by $\mathcal{I}_f \equiv \{f(1), f(2), ..., f(k)\} \subset [M] \times [M]$.

AMIS Formalization

Problem Setting

In particular for $f \in \mathbf{F}_k$, let

$$X_f \equiv (X_{f(1)}, X_{f(2)}, .., X_{f(k)}), \tag{4}$$

denote the sensed random vector with values in $\mathcal{A}^k$ and let

$$\hat{X}_f \equiv (X_{i,j} : (i,j) \in [M] \times [M] \setminus \{f(1), f(2), .., f(k)\}) \tag{5}$$

denote the non-sensed random vector with values in $\mathcal{A}^{M^2-k}$. In this context, given some specific sensed values $\bar{x} = (x_1, .., x_k) \in \mathcal{A}^k$, the complexity of simulating the non-sensed position $\hat{X}_f$ is given by

$$\begin{aligned} H(\hat{X}_f | X_f = \bar{x}) &= \mathcal{H}(\mu_{\hat{X}_f|X_f}(\cdot|\bar{x})) \\ &= - \sum_{\bar{y}=(y_1,..,y_{M^2-k}) \in \mathcal{A}^{M^2-k}} \mu_{\hat{X}_f|X_f}(\bar{y}|\bar{x}) \cdot \log_2 \mu_{\hat{X}_f|X_f}(\bar{y}|\bar{x}). \end{aligned} \tag{6}$$

Case Study

The proposed model is the one dimensional array described in Fig. 7.

A Binary Markov Chain

X_1	X_2	$\dots$	X_{i-1}	X_i	X_{i+1}	$\dots$	X_{N-1}	X_N

Figure 7: *1-D* regionalized variable.

In the markovian scenario, the probabilistic model exhibits the following spatial dependence: given the present, the future is independent of the past. For the proposed setting, see Fig. 7, a past-present-future sorting has been imposed from location 1 to the location N.

Conditioning to any arbitrary subset of states

Arbitrary subset of states

$\dots$	X_{b_1}	$\dots$	X_{b_2}	$\dots$	b_B	$\dots$	X_i	$\dots$	X_{a_1}	$\dots$	X_{a_2}	$\dots$	X_{a_A}	$\dots$

Figure 8: Measured Variables in separated past and future subsets.

The target entropy of the conditional distribution for the state X_i, can be defined as:

$$
\begin{aligned}
H(X_i \mid \left\{X_{b_j}\right\} &= \left\{x_{b_j}\right\}, \{X_{a_j}\} = \{x_{a_j}\}) \\
&= -\sum_{x_i \in \mathcal{A}} p(x_i \mid \left\{x_{b_j}\right\}, \{x_{a_j}\}) \cdot log\ p(x_i \mid \left\{x_{b_j}\right\}, \{x_{a_j}\}) \\
\text{by Eq. (??)} &= -\sum_{x_i \in \mathcal{A}} p(x_i \mid x_{b_B}, x_{a_1}) \cdot log\ p(x_i \mid x_{b_B}, x_{a_1}) \\
&= H(X_i \mid X_{b_B} = x_{b_B}, X_{a_1} = x_{a_1})
\end{aligned}
\tag{7}
$$

Entropy for arbitrary subset of states

$$H(X_i \mid X_{b_B}, X_{a_1}) = -\sum_{x_{b_B} \in \mathcal{A}} \sum_{x_i \in \mathcal{A}} \sum_{x_{a_1} \in \mathcal{A}} [(\\ p(X_{b_B} = x_{b_B}) \cdot p(X_i = x_i \mid X_{b_B} = x_{b_B}) \cdot p(X_{a_1} = x_{a_1} \mid X_i = x_i) \\ ) \cdot log \left(\frac{p(x_{a_1} \mid x_i) \cdot p(x_i \mid x_{b_B})}{p(x_{a_1} \mid x_{b_B})} \right)] \tag{8}$$

The Iterative Sequential Rule SMIS

Sequential (non-adaptive) maximum information scheme (SMIS) is proposed as an iterative solution based on the principle of one-step-ahead sensing.

Induced Iterative Rule

Iterating this inductive rule, the k-measurement (after solving (i_1^*, j_1^*), (i_2^*, j_2^*) , .. , (i_{k-1}^*, j_{k-1}^*)) is the solution of SMIS at stage k, that is,

$$(i_k^*, j_k^*) = \arg \max_{(i,j)\in[M]\times[M]\setminus\left\{(i_l^*, j_l^*): l=1,..,k-1\right\}} H(X_{i,j}|X_{i_1^*, j_1^*}, .., X_{i_{k-1}^*, j_{k-1}^*}). \quad (9)$$

Therefore, with this sequence of optimal positions $\left\{(i_l^*, j_l^*) : \ l = 1, .., k\right\}$, for every $k \in \left\{1, .., M^2\right\}$, the sequential rule $\tilde{f}_k^* \in \mathbf{F}_k$ can be constructed by

$$\tilde{f}_k^*(1) = (i_1^*, j_1^*), \tilde{f}_k^*(2) = (i_2^*, j_2^*), ..., \text{ and } \tilde{f}_k^*(k) = (i_k^*, j_k^*). \quad (10)$$

The Adaptive Sensing Problem AMIS

The adaptive maximum information sampling (AMIS) is introduced as an adaptive sensing variation for the sequential strategy elaborated in Sect. **??**.

Formally,

Instead of considering the information gain in average, with respect to the statistics of the random vector $(X_{i_1^*,j_1^*}, X_{i_2^*,j_2^*}, \dots, X_{i_{k-1}^*,j_{k-1}^*})$ in (9), an adaptive strategy can condition on the specific values previously measured at the $k-1$ positions. Then, assuming access to the "*true data*" $(x_1,..,x_{k-1}) \in \mathcal{A}^{k-1}$ of the image at the positions $(i_1^a, j_1^a),..,(i_{k-1}^a, j_{k-1}^a)$, the next position is the solution of the AMIS approach, given by

$$(i_k^a(x_1,..,x_{k-1}), j_k^a(x_1,..,x_{k-1})) = \arg\max_{(i,j)\in[M]\times[M]\setminus\left\{(i_l^a,j_l^a):l=1,..,k-1\right\}} H(X_{i,j}|X_{i_1^a,j_1^a} = x_1,..,X_{i_{k-1}^a,j_{k-1}^a} = x_{k-1}). \tag{11}$$

AMIS approach

The Adaptive Sensing Problem

The next position to be sampled is the solution of the AMIS approach, given by

$$(i_k^a(x_1,..,x_{k-1}), j_k^a(x_1,..,x_{k-1})) = \underset{(i,j)\in[M]\times[M]\setminus\{(i_l^a,j_l^a):l=1,..,k-1\}}{\arg\max} H(X_{i,j}|X_{i_1^a,j_1^a}=x_1,..,X_{i_{k-1}^a,j_{k-1}^a}=x_{k-1}). \tag{12}$$

The solution in (12) is a function of the following set of marginal conditional distributions in $\mathcal{P}(\mathcal{A})$

$$\left\{\mu_{X_{i,j}|X_{i_1^a,j_1^a},..,X_{i_{k-1}^a,j_{k-1}^a}}(\cdot|x_1,..,x_{k-1}) : (i,j) \text{ non-sensed at iteration } k-1\right\} \tag{13}$$

and, consequently, of the measured data $(x_1,..,x_{k-1})$.

AMIS approach

Information Gain

The reduction of uncertainty or the information gained to resolve the field, attributed to the adaptive decision rule that solves (12), given the previous positions (i_1^a, j_1^a) , .. , (i_{k-1}^a, j_{k-1}^a) and their data $(x_1, .., x_{k-1})$, at the stage k is denoted and expressed by

$$\underbrace{I(\tilde{f}_k^a, (x_1, .., x_{k-1}))}_{\text{information gain}} \equiv \underbrace{H(\hat{X}_{\tilde{f}_{k-1}^a} | X_{\tilde{f}_{k-1}^a} = (x_1, .., x_{k-1}))}_{\text{entropy of the prior model at stage } k-1} -$$

$$\underbrace{H(\hat{X}_{\tilde{f}_{k-1}^a} | X_{i_k^a, j_k^a}, X_{\tilde{f}_{k-1}^a} = (x_1, .., x_{k-1}))}_{\text{minimum posterior entropy when selecting } (i_k^a, j_k^a) \text{ at stage } k-1},$$

$$= H(X_{i_k^a, j_k^a} | X_{i_1^a, j_1^a} = x_1, .., X_{i_{k-1}^a, j_{k-1}^a} = x_{k-1}) \geq 0, \quad (14)$$

where (i_k^a, j_k^a) is a short-hand for the solution of (12) that is a function of $(x_1, .., x_{k-1})$. In fact, by using information quantities, $I(\tilde{f}_k^a, (x_1, .., x_{k-1}))$ is precisely the mutual information between $\hat{X}_{\tilde{f}_{k-1}^a}$ (the vector of non-sensed positions at the stage $k-1$) and $X_{i_k^a, j_k^a}$ (solution of (12)) conditioned by $X_{\tilde{f}_{k-1}^a} = (x_1, .., x_{k-1})$.

AMIS approach

Diagram to solve the adaptive rule in Eq. (12)

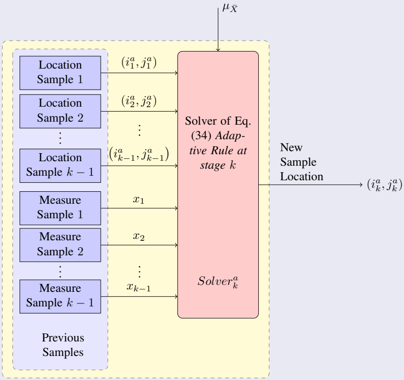


Figure 9: Diagram of the inputs and statistical information (model) needed to solve the adaptive rule in Eq. (12)

Empirical Analysis

As a performance indicator, $C(f_k)$ is introduced as the information of f_k to resolve the field normalized by the entropy of the entire field $H(\bar{X})$.

In particular,

$$C(f_k) \equiv \frac{I(f_k)}{H(\bar{X})} = \frac{\sum_{i=2}^{k-1} H(X_{f(i)}|X_{f(i-1)}) + H(X_{f(1)})}{\sum_{i=2}^{N} H(X_i \mid X_{i-1}) + H(X_1)} \in [0,1]. \tag{15}$$

To provide insight note that: $C(f_k) = 0$ means that the k-measurements produce no reduction in uncertainty, and $C(f_k) = 1$ means that there is no remaining uncertainty to be resolved after taking the k-measurements, that is, $H(\tilde{X}_{f_k}|\bar{X}_{f_k}) = 0$.

Conditional Entropies after sampling with SMIS

For this analysis, $N = 1000$ and a symmetric stochastic matrix $\mathbf{A}$ with transition probability $\beta = 0{,}01$ and X_1 uniformly distributed in $\{0, 1\}$ are considered.

Conditional entropies for a binary Markov chain

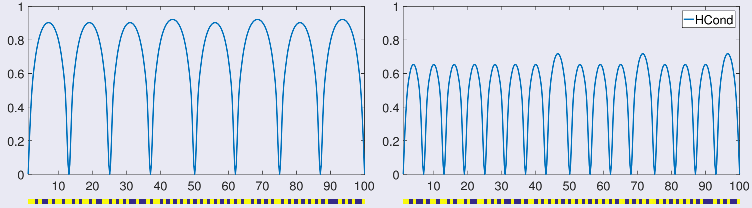


Figure 10: Conditional entropies given the sensed positions for a binary Markov chain. Left: after the first 10 samples. Right: after 18 samples. $\beta = 0{,}01$. Under the curves the actual realizations of the Markov chains are presented

Uncertainty Reduction for SMIS

.

Uncertainty Reduction

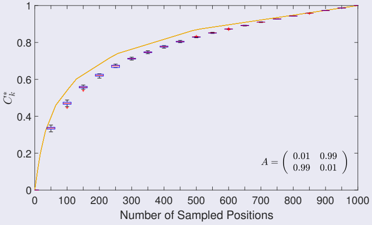


Figure 11: Uncertainty reduction for 20 realizations of the Markov chain. Continuous curves: SMIS; boxplots: random sampling. Left: $\beta = 0{,}01$, Right: $\beta = 0{,}8$

Sampling with AMIS

Conditional entropies for a binary Markov chain

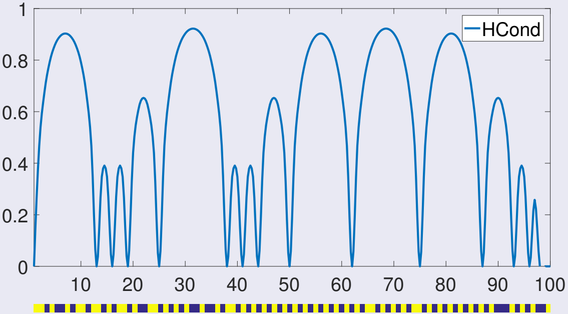


Figure 12: Distribution of the conditional entropy of non-sensed locations given the sensed pixels at $k = 18$, using AMIS. $\beta = 0{,}01$. Under the curve, the actual realization of the binary Markov chain is presented

Comparison SMIS and AMIS

SMIS vs AMIS

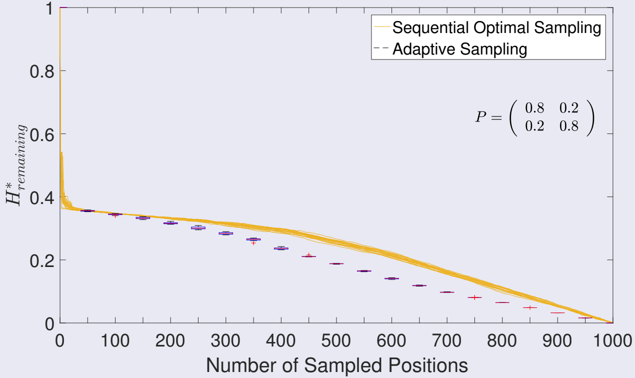


Figure 13: Remaining conditional entropy by considering the previous sampled locations and its measurements. Symmetric transition matrix $(\beta = 0{,}2)$

Estimation Error

AMIS vs Random Sampling

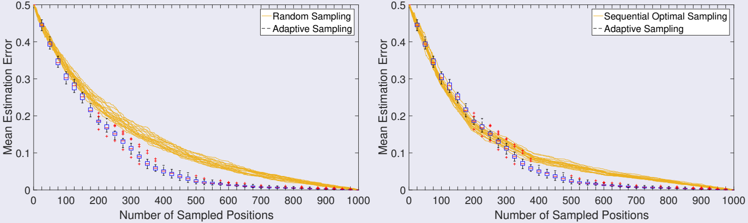


Figure 14: Estimation error considering the previous sampled locations and its measurements. Symmetric transition matrix with $\beta = 0{,}1$. Left: Random Sampling vs AMIS Method, Right: SMIS vs AMIS

RAMIS Proposal

The preferential sampling solution proposed in this work is a combination between the pure *AMIS* in (12) and a non-adaptive rule reminiscing of the *SMIS* strategy in (9) under a Markov assumption.

RAMIS

Let $S_{k-1} = \left\{(i_l^a, j_l^a) : l = 1, .., k-1\right\}$ denote the collection of the sampled locations obtained by the proposed adaptive sampling strategy at the stage $k-1$ of the algorithm. Here, $X_{S_{k-1}} = (X_{i_1^a, j_1^a}, .., X_{i_{k-1}^a, j_{k-1}^a})$ corresponds to the sensed random vector indexed by S_{k-1}, and $x_{S_{k-1}} = (x_1, .., x_{k-1})$ denotes the measurements taken at S_{k-1} in $\mathcal{A}^{k-1}$. Thus, the *regularized AMIS* rule (*RAMIS*) for stage k is the solution of

$$(\hat{i}_k^a(\alpha, x_{S_{k-1}}), \hat{j}_k^a(\alpha, x_{S_{k-1}})) = \underset{(i,j)\in[M]\times[M]\setminus S_{k-1}}{\arg\max} \alpha \cdot H(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}}) + (1-\alpha) \cdot D((i,j), S_{k-1}). \quad (16)$$

Application to binary channels

Database

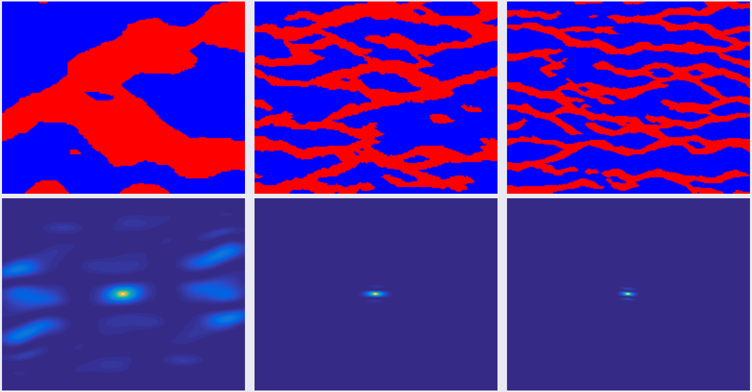


Figure 15: Proposed training images. Top: example of TIs, Bottom: mean MI maps estimated from 200 realizations. From left to right: Models SC_1, MC_1, MC_2. Color maps, Top: Red is channel presence; Bottom: Linear from blue (low MI) to bright yellow (max. MI)

Application to binary channels

For performance evaluation two metrics were considered: resolvability and simulation error.

Metrics

The resolvability of f is the average conditional entropy over the non-sensed positions, given by

$$\begin{aligned} \mathcal{R}(f, x_{S_k}) &\equiv average(H(X_{i,j}|X_{S_k} = x_{S_k})_{(i,j)\in[M]\times[M]\setminus S_k}) \\ &= \frac{1}{M^2 - |S_k|} \sum_{(i,j)\in[M]\times[M]\setminus S_k} H(X_{i,j}|X_{S_k} = x_{S_k}) \end{aligned} \tag{17}$$

Application to binary channels

For performance evaluation two metrics were considered: resolvability and simulation error.

Metrics

If $(x_{i,j})_{i,j}$ is the true image, then the simulation error induced from f is the average error over the non-sensed positions of the simulations given by

$$\begin{aligned}
&\mathcal{E}(f,(x_{i,j})_{i,j}) \equiv \\
&\frac{1}{M^2-|S_{k-1}|}\sum_{(i,j)\in[M]\times[M]\setminus S_k}\mathbb{E}_{X_{i,j}}\left\{(x_{i,j}-X_{i,j})^2\,|X_{S_k}=x_{S_k}\right\} \\
&=\frac{1}{M^2-|S_{k-1}|}\sum_{(i,j)\in[M]\times[M]\setminus S_k}\mathbb{E}_{X_{i,j}}\left\{\mathbf{1}_{X_{i,j}\neq x_{i,j}}|X_{S_k}=x_{S_k}\right\} \\
&=\frac{1}{M^2-|S_{k-1}|}\sum_{(i,j)\in[M]\times[M]\setminus S_k}\underbrace{\mathbb{P}\left\{X_{i,j}\neq x_{i,j}|X_{S_k}=x_{S_k}\right\}}_{\text{Conditional Error Probability}}.
\end{aligned} \tag{18}$$

From (18), $\mathcal{E}(f,(x_{i,j})_{i,j})$ corresponds to the average frequency of error in detecting the true non-sensed value with the values simulated from MPS, over the non-sensed locations.

Selecting the Regularization Parameter α

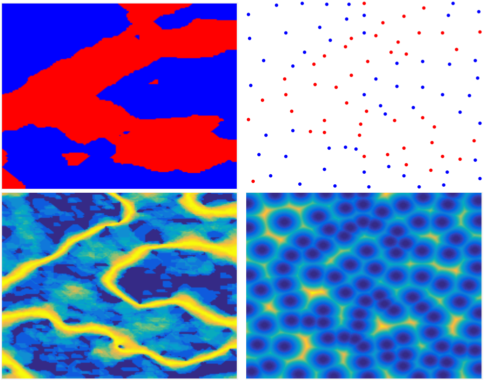


Figure 16: Estimated maps for the model SC_1 at $k = 100$. Upper Row, Left: Reference Image, Right: Sampled Locations. Lower Row, Left: Entropy Map ($\alpha = 1$), Right: Distance Map ($\alpha = 0$). Color maps, linear from blue to bright yellow (from low to high entropy or distances)

Selecting the Regularization Parameter α

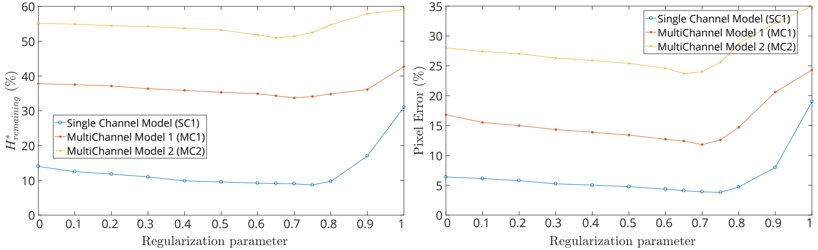


Figure 17: Performance of non-sensed positions under RAMIS as a function of α, after 500 samples. Left: Resolvability, Right: Mean error. Average curves for 50 independent train-test sets

Performance

Table 1: Global Performance Improvement after sampling $1{,}25\,\%$ of positions (500 Samples in images of size 200 by 200 pixels). Here, the outcome for α providing the best performance for each model is presented.

Model	Mean Metric	Reference Performance (%)	Optimal Performance (%)	Absolute Improvement (%)	Relative Improvement (%)
Model SC1 ($\alpha : 0{,}75$)	Entropy	14.0	8.7	5.3	37.85
	Pixel Error	6.4	3.8	2.6	40.63
Model MC1 ($\alpha : 0{,}70$)	Entropy	37.8	33.7	4.1	10.85
	Pixel Error	16.8	11.8	5.0	29.76
Model MC2 ($\alpha : 0{,}65$)	Entropy	55.1	51.0	4.1	7.44
	Pixel Error	28.0	23.7	4.3	15.3

Performance

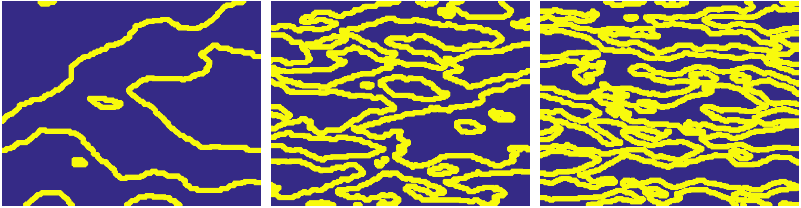
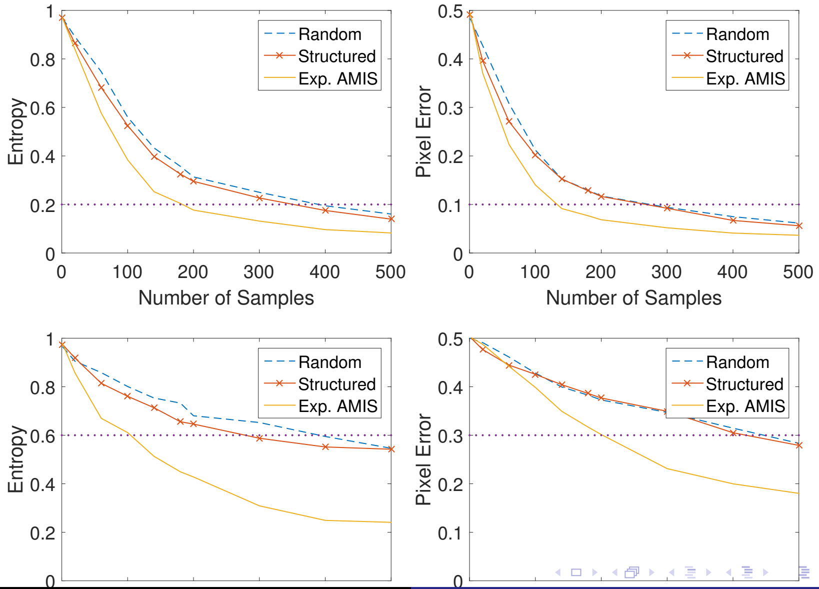


Figure 18: Example of the masks used to define transitions in channelized images. 5 pixels around the transitions are considered. From left to right: models SC_1, MC_1, and MC_2. Color maps, solid yellow is the mask

Performance



Performance Single Channel

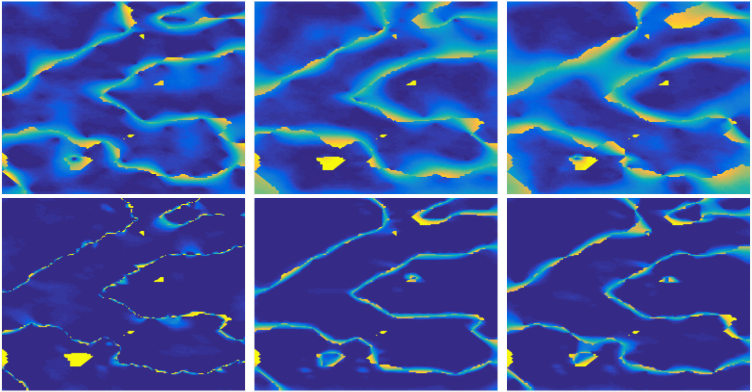


Figure 20: Remaining Entropy maps for model SC_1 using 600 realizations of the sampling process. Top: maps using the first 100 samples. Bottom: maps using the first 500 samples. Left to right: RAMIS, quasi-regular, and random sampling. Color maps, linear from blue (low remaining entropy) to bright yellow (high remaining entropy)

Adaptive Ore-waste discrimination

Ore-waste discrimination with adaptive sampling strategy

Diagram

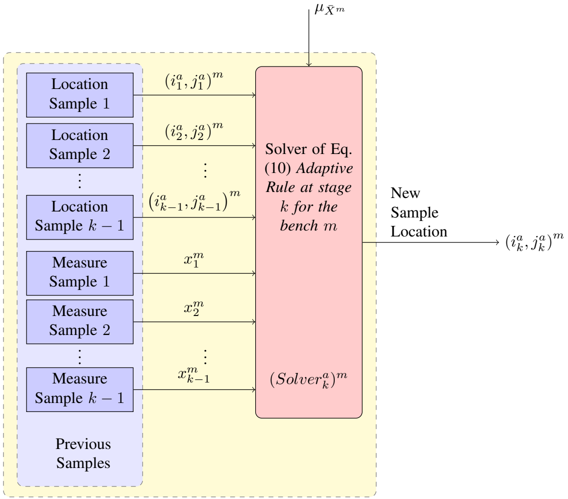


Figure 21: Schematic diagram of the sampling rule (**??**).

Case Studies

The proposed methodology has been applied to three different cases, two of them coming from the same ore deposit. The corresponding databases consist of drill-hole composites widely spaced and denser blast-hole samples, which are used to validate the sampling strategy. The two projects correspond to massive porphyry copper deposits that are currently under operation

Case Studies

The basic statistical information used to build the case studies is summarized in Table 2 and Table 3.

Table 2: Summary statistics.

	Case Study 1		Case Study 2		Case Study 3	
	Drill-hole Samples	Blast-hole Samples	Drill-hole Samples	Blast-hole Samples	Drill-hole Samples	Blast-hole Samples
Count	2045	19752	747	95815	2368	158772
Mean	1.07	1.18	0.34	0.42	0.57	0.48
Std. dev.	0.67	0.78	0.47	0.56	0.55	0.58
Minimum	0.13	0.01	0.01	0.00	0.00	0.00
Maximum	7.24	9.90	4.04	35.00	4.04	35.00

Case Studies

Table 3: Case study coordinates. Elevations represent the centers of the considered benches.

	Case Study 1		Case Study 2		Case Study 3	
	Min	Max	Min	Max	Min	Max
East	24550	24730	72200	72550	72600	72900
North	25100	25550	83100	83500	83100	83600
Elevation	3860	3940	2405	2455	2415	2465

Case Studies

The statistical distributions of blast-holes grades present in the analyzed case studies are described in Fig. 22, along with their basic statistics.

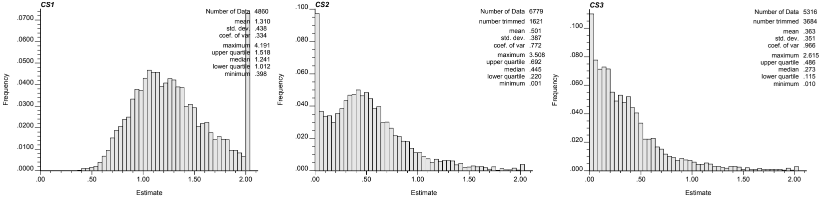


Figure 22: Grade mineral distributions and basic statistics for the available blast-holes. From left to right: CS1, CS2, CS3.

Case Study I: Drill-hole samples data

In order to illustrate the density of available information for every single bench, Figs. 23 and 24 show the drill-hole composites and blast-hole samples for the first case study. The block model estimated by ordinary kriging is displayed for these data in Fig. 25.

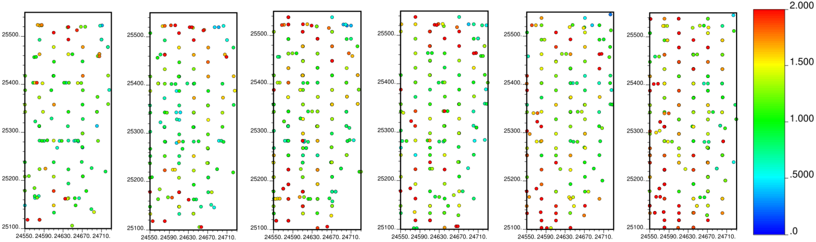


Figure 23: Drill-hole samples data for case study 1. From left to right: Benches 1-6. Colormap denotes the grade of *Cu*.

Case Study I: Blast-hole data

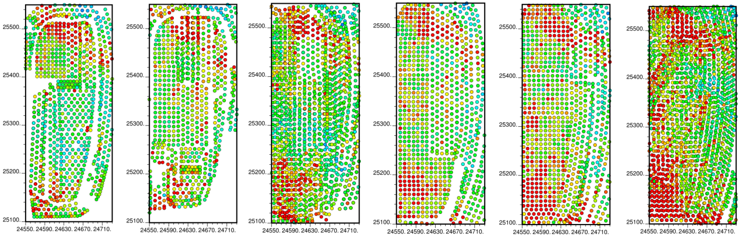


Figure 24: Blast-hole data for case study 1. From left to right: Benches 1-6. Colormap: the same as in Fig. 23.

Case Study I: Ground truth Estimation

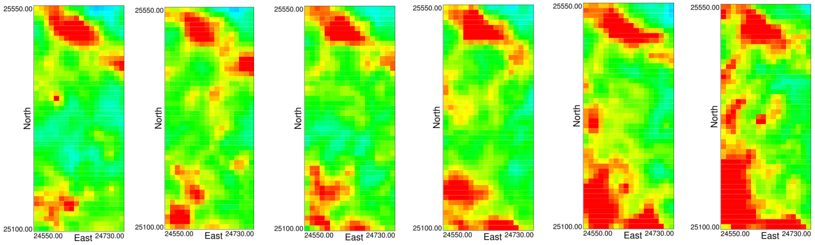


Figure 25: Ground truth estimated from drill-holes and blast-holes samples for case study 1. From left to right: Benches 1-6. Colormap: the same as in Fig. 23.

Case Study I: Samples from RAMIS

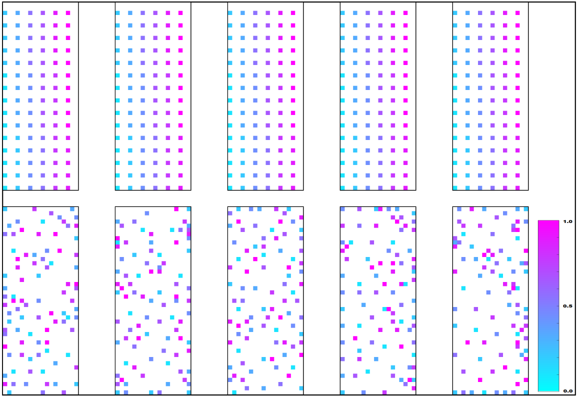


Figure 26: Samples for Case Study 1. From left to right: Benches *2-6*. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade $1{,}012\,\%$. Colormap: Describe batch of samples in the order of the performed sampling.

Case Study I: Reconstructions

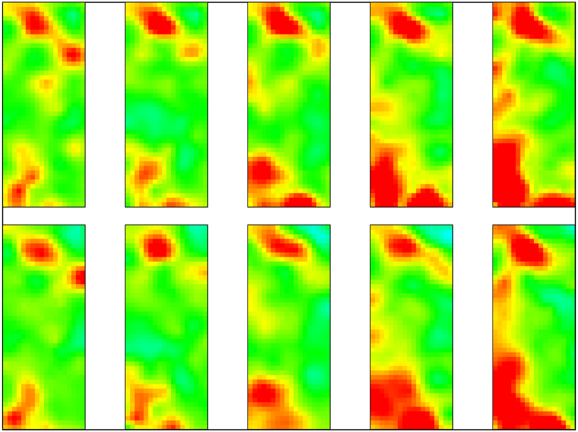
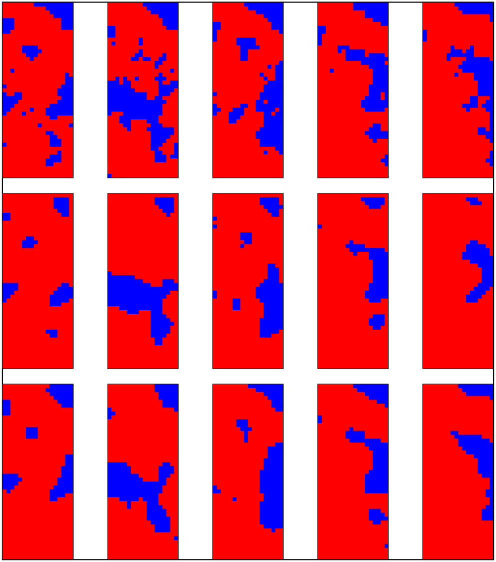


Figure 27: Estimated grade for Case Study 1. From left to right: Benches *2-6*. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade $1{,}012\,\%$. Colormap: the same as in Fig. 23.

Case Study I: Estimated grade



Case Study I: Confusion Matrix

Figure 29: Confusion Matrix for Case Study 1. From left to right: Benches *2-6*. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade $1{,}012\,\%$.

Case Study I: Binary Image Inference Performance

Table 4: Performance Error Summary for Case Study 1.

	Case Study 1					
	Cutoff grade 1,102		Cutoff grade 1,241		Cutoff grade 1,518	
	STR	ADA	*STR*	ADA	*STR*	ADA
Bench 2	0.112	**0.069**	0.133	**0.114**	0.059	**0.055**
Bench 3	0.138	**0.106**	0.104	**0.102**	0.066	**0.064**
Bench 4	0.109	**0.082**	0.091	**0.086**	0.053	**0.051**
Bench 5	0.084	**0.048**	0.086	**0.083**	0.090	**0.083**
Bench 6	0.111	**0.060**	0.127	**0.108**	0.109	**0.097**

Case Study II: Binary Image Inference Performance

Table 5: Performance Error Summary for Case Study 2.

	Case Study 2					
	Cutoff grade $0{,}220$		Cutoff grade $0{,}445$		Cutoff grade $0{,}692$	
	STR	ADA	*STR*	ADA	*STR*	ADA
Bench 2	0.048	**0.043**	0.100	**0.096**	0.112	**0.080**
Bench 3	0.039	**0.032**	0.109	**0.097**	0.091	**0.068**
Bench 4	0.037	**0.034**	0.078	**0.066**	0.057	**0.045**
Bench 5	0.055	**0.038**	0.036	**0.024**	0.036	**0.026**
Bench 6	0.031	**0.015**	0.025	**0.010**	0.013	**0.010**

Case Study III: Binary Image Inference Performance

Table 6: Performance Error Summary for Case Study 3.

	Case Study 3					
	Cutoff grade $0{,}115$		Cutoff grade $0{,}273$		Cutoff grade $0{,}486$	
	STR	ADA	*STR*	ADA	*STR*	ADA Bench 2
0.067	**0.048**	0.060	**0.039**	0.067	**0.054**	
Bench 3	0.039	**0.031**	0.057	**0.030**	0.030	**0.014**
Bench 4	0.049	**0.029**	0.052	**0.033**	0.054	**0.049**
Bench 5	0.051	**0.037**	0.053	**0.041**	0.053	**0.034**
Bench 6	0.037	**0.021**	0.047	**0.035**	0.038	**0.034**

Case Study I: Economical Profit Estimation

Table 7: Economical Profit Estimation for Case Study 1. In MM $US\$$.

	Case Study 1					
	Cutoff grade $1{,}102$		Cutoff grade $1{,}241$		Cutoff grade $1{,}518$	
	STR	ADA	*STR*	ADA	*STR*	ADA
Bench 2	33.574	**34.146**	13.483	**13.726**	**2.675**	1.737
Bench 3	**28.581**	28.462	**10.419**	9.439	1.284	**2.520**
Bench 4	33.562	**34.150**	15.423	**16.114**	**6.295**	5.792
Bench 5	41.065	**41.590**	23.272	**23.814**	11.174	**11.502**
Bench 6	46.401	**47.135**	27.623	**28.165**	**16.420**	15.427
Global	183.183	**185.483**	90.221	**91.257**	**37.849**	36.977

Case Study II: Economical Profit Estimation

Table 8: Economical Profit Estimation for Case Study 2. In MM $US\$$.

	Case Study 2					
	Cutoff grade $0{,}220$		Cutoff grade $0{,}445$		Cutoff grade $0{,}692$	
	STR	ADA	*STR*	ADA	*STR*	ADA
Bench 2	49.281	**49.320**	23.729	**23.992**	6.096	**6.704**
Bench 3	39.309	**39.458**	**16.727**	16.362	-5.363	**-4.555**
Bench 4	**47.299**	47.262	23.257	**23.646**	8.957	**9.425**
Bench 5	36.460	**36.703**	**19.278**	19.181	10.392	**10.714**
Bench 6	9.790	**9.994**	1.145	**1.330**	-4.058	**-3.937**
Global	182.139	**182.735**	84.137	**84.512**	16.023	**18.351**

Case Study III: Economical Profit Estimation

Table 9: Economical Profit Estimation for Case Study 3. In MM $US\$$.

	Case Study 3					
	Cutoff grade $0{,}115$		Cutoff grade $0{,}273$		Cutoff grade $0{,}486$	
	STR	ADA	*STR*	ADA	*STR*	ADA
Bench 2	17.685	**17.974**	**17.382**	17.371	**3.059**	2.419
Bench 3	11.577	**11.674**	10.136	**11.130**	-2.159	**-1.943**
Bench 4	10.651	**10.904**	9.772	**10.213**	**-2.440**	-2.553
Bench 5	14.629	**14.859**	**14.111**	13.993	1.192	**1.725**
Bench 6	16.038	**16.288**	**15.316**	15.262	3.136	**3.964**
Global	70.580	**71.699**	66.717	**67.969**	2.787	**3.611**

Conclusions

- The role of preferential sampling has been systematically addressed for the task of geological facies recovery using multiple-point simulation (*MPS*) and for the problem of short-term planning in mining.
- In the context of facies recovery using simulations, the task of optimal sampling is formalized and addressed using a maximum information extraction criterion. This sampling principle has the ability to locate samples adaptively on the positions that extract maximum information for the objective of resolving the underlying field.
- A formal justification is provided for adopting this information-driven sampling criterion as well as concrete ways of implementing this principle in practice.

Conclusions

- The proposed sampling strategy has been adapted to the problem of short-term planning for the task of classifying blocks to be processed as waste or ore in the production stage of a mining project.
- The proposed methodology takes advantage of the information available from the previously sampled locations, allowing to improve the performance as compared with some of the classical non-adaptive sampling schemes used for advanced drilling tasks.
- It is important to emphasize that no previous work have addressed the optimal sensing problem covered in this thesis for characterization of geological fields in the context of *MPS*.

Future work

Future work:

- The applied principles can be extended to the characterization and recovery of other geological signals with spatial structure in under sampling contexts.
- There are many directions where this idea could be applied and it is an interesting direction of future research to explore the full potential of this framework. On the specifics, it would be interesting to apply the proposed strategy to scenarios with multiple categories and to use techniques for geostatistical continuous simulation to extend the proposed methodology to continuous variables.
- Study alternative geostatistical simulation tools that could provide more effective estimations of the multi-point patterns.
- The results are applicable to a wide range of disciplines where similar sampling problems need to be faced, included but not limited to design of communication networks, optimal integration and communication of swarms of robots and drones, remote sensing.

[1] Bangerth, W., Klie, H., Matossian, V., Parashar, M., and Wheeler, M.
An autonomic reservoir framework for the stochastic optimization of well placement.
Cluster Computing 8 (2005), 255–269.

[2] Bangerth, W., Klie, H., Wheeler, M. F., Stoffa, P., and Sen, M.
On optimization algorithms for the reservoir oil well placement problem.
Comp. Geosciences 10 (2006), 303–319.

[3] Bangerth, W., Klie, H., Wheeler, M. F., Stoffa, P. L., and Sen, M. K.
On optimization algorithms for the reservoir oil well placement problem.
COMP. GEOSC (2004).

[4] Cover, T., and Thomas, J.
Elements of Information Theory.
Elements of Information Theory. Wiley, 2006.

[5] Cross, G. R., and Jain, A. K.
Markov random field texture models.
Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-5, 1 (Jan 1983), 25–39.

[6] Guestrin, C., Krause, A., and Singh, A.
Near-optimal sensor placements in Gaussian processes.
In *International Conference on Machine Learning (ICML)* (August 2005).

[7] Krause, A.
Optimizing Sensing: Theory and Applications.
PhD thesis, Carnegie Mellon University, December 2008.

[8] Krause, A., and Guestrin, C.
Near-optimal observation selection using submodular functions.
In *National Conference on Artificial Intelligence (AAAI), Nectar track* (July 2007).

[9] KRAUSE, A., GUESTRIN, C., GUPTA, A., AND KLEINBERG, J.
Near-optimal sensor placements: Maximizing information while minimizing communication cost.
In *Proc. of Information Processing in Sensor Networks (IPSN)* (2006).

[10] KRAUSE, A., GUESTRIN, C., GUPTA, A., AND KLEINBERG, J.
Robust sensor placements at informative and communication-efficient locations.
ACM Transactions on Sensor Networks (TOSN) 7, 4 (February 2011).

[11] KRAUSE, A., LESKOVEC, J., GUESTRIN, C., VANBRIESEN, J., AND FALOUTSOS, C.
Efficient sensor placement optimization for securing large water distribution networks.
Journal of Water Resources Planning and Management 134, 6 (November 2008), 516–526.

[12] KRAUSE, A., RAJAGOPAL, R., GUPTA, A., AND GUESTRIN, C.
Simultaneous placement and scheduling of sensors.
In *Proc. ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)* (2009).

[13] Krause, A., Singh, A., and Guestrin, C.
Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies.
Journal of Machine Learning Research (JMLR) 9 (February 2008), 235–284.

[14] Oliver, D. S., Reynolds, A. C., and Liu, N.
Inverse Theory for Petroleum Reservoir Characterization and History Matching.
Cambridge University Press, Cambridge, 2008.

[15] Onwunalu, J., and Durlofsky, L.
Application of a particle swarm optimization algorithm for determining optimum well location and type.
Computational Geosciences (2009).

[16] Wellmann, J. F.
Information theory for correlation analysis and estimation of uncertainty reduction in maps and models.
Entropy 15, 4 (2013), 1464–1485.

Thanks!

With an everlasting love, in memory of

Elba del Carmen Bustos Alarcón
and *Hugo Felipe Santibáñez Salazar*
and *Julia Teresa Salazar Fuentealba*
and *Blanca Flor Leal Tapia*
and *Mercedes Tapia Tapia*
and *C. P. R. E.*