



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

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

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN INGENIERÍA ELÉCTRICA

FELIPE ANDRÉS SANTIBÁÑEZ LEAL

PROFESOR GUÍA:
PH.D. JORGE SILVA SANCHEZ
PROFESOR CO- GUÍA:
PH.D. JULIÁN ORTIZ CABRERA

MIEMBROS DE LA COMISIÓN:
JAVIER RUIZ DEL SOLAR
RONNY VALLEJOS ARRIAGADA
MATIAS ZAÑARTU SALAS

Este trabajo ha sido parcialmente financiado por CONICYT PHD Fellowship 2013, 21130890; BNI Postdoctoral Bridge Fellowship 2017; The Information and Decision Systems Group (IDS) (Dep. of Electrical Engineering, University of Chile); The Advanced Laboratory for Geostatistical Supercomputing (ALGES) (Advanced Mining Technology Center (AMTC) CONICYT Project AFB180004, Dep. of Mining Engineering, University of Chile).

SANTIAGO DE CHILE
2019

RESUMEN DE LA TESIS PARA OPTAR
AL TÍTULO DE DOCTOR EN INGENIERÍA ELÉCTRICA
POR: FELIPE ANDRÉS SANTIBÁÑEZ LEAL
FECHA: DICIEMBRE 2019
PROF. GUÍA: PH.D. JORGE SILVA SANCHEZ PH.D. JULIÁN ORTIZ CABRERA

ESTRATEGIA DE MUESTREO BASADO EN TEORÍA DE INFORMACIÓN PARA LA
RECUPERACIÓN DE IMÁGENES GEOLOGICAS: MODELAMIENTO, ANÁLISIS, E
IMPLEMENTACIÓN

Las herramientas *Geoestadísticas* se han convertido en el estándar para la caracterización de la distribución espacial de estructuras geológicas subterráneas. Sin embargo, el problema de recuperación de imágenes en regímenes de bajas tasas de adquisición continúa siendo una tarea desafiante. En la última década, se han desarrollado diversos métodos alternativos para el diseño experimental en estos regímenes, proveyendo nuevas perspectivas para alcanzar mejoras en el desempeño en tareas de reconstrucción y caracterización de imágenes geológicas. Con base en estos logros, un nuevo desafío recae en incorporar herramientas del *estado del arte* en procesamiento de señales y modelamiento estocástico para mejorar este tipo de problemas de inferencia. Esta tesis ha propuesto un estudio profundo de problemas inversos a bajas tasas de muestro con un fuerte enfoque en Geociencias y, en particular, en la reconstrucción de canales binarios de permeabilidad y tareas de control de ley en el contexto de planificación minera de corto plazo.

En este trabajo, la formulación y análisis experimental del problema de *Posicionamiento Óptimo de Sensores (OSP)* han sido investigados en el contexto de modelos categóricos *2-D* con dependencia espacial. En minería, este problema intenta encontrar la mejor forma de distribuir las mediciones para optimizar el muestreo/localización de recursos en áreas de exploración y explotación minera. Este trabajo ha apuntado a la formalización del problema *OSP* para un número dado de mediciones disponibles. La caracterización de la incertidumbre es una pieza central de esta formalización. En particular, *OSP* ha sido abordado desde la perspectiva de minimizar la incertidumbre remanente y mediante la implementación de algoritmos secuenciales capaces de estimar dicha incertidumbre. El uso de conceptos de teoría de información (*IT*) como *entropía condicional* ha sido considerado para caracterizar la incertidumbre relacionada con modelos geológicos condicionados a la adquisición de datos, y su aplicación en una estrategia de muestreo preferencial orientada a mejorar la inferencia geoestadística en regímenes de bajas tasas de adquisición. La conjetura ha sido que las locaciones basadas en *IT-OSP* se distribuyen en la zona de transición de campos aleatorios categóricos, mejorando el desempeño en tareas de recuperación de imágenes comparado con esquemas de muestro no adaptivos clásicos.

A nivel experimental, un algoritmo secuencial regularizado ha sido implementado para aproximar el muestro *IT-OSP* para mostrar las ventajas de este principio. El enfoque propuesto proporciona realizaciones basadas en simulaciones multipunto con variabilidad reducida para modelos de facies geológicas categóricas en regímenes de muestreo críticos. Finalmente, el desempeño de los procesos de inferencia bajo las estrategias de muestreo propuestas ha sido evaluado en escenarios prácticos realistas para las tareas de control de ley en el contexto de planificación de corto plazo.

SUMMARY OF THE THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR EN INGENIERÍA ELÉCTRICA
BY: FELIPE ANDRÉS SANTIBÁÑEZ LEAL
DATE: DECEMBER 2019
ADVISOR: PH.D. JORGE SILVA SANCHEZ PH.D. JULIÁN ORTIZ CABRERA

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

Geostatistical tools have become the standard for characterizing the spatial distribution of geological subsurface structures. However, the problem of image recovery for regimes with low acquisition rates still poses a complex issue. In the last decade, several alternative methods for experimental design at low sampling rates has been developed providing insights into the use of additional prior information to achieve better performance in the reconstruction and characterization of geological images. Based on these achievements, a new challenge is to incorporate tools from the *state of art* in signal processing and stochastic modeling to improve this kind of inference problems. This thesis proposed a comprehensive study of inverse problems at low sampling rates with strong focus on Geosciences and, in particular, for the reconstruction of binary permeability channels and for grade control tasks in short term planning.

In this work, the formulation and experimental analysis of the *Optimal Sensor Placement (OSP)* problem has been investigated in the context of categorical *2-D* models with spatial dependence. In the mining exploration and production area, this problem attempts to find the best way of distributing measurements (or samples) to optimize sensing/locating resources in areas of mining and drilling. This work aims at formalizing the *OSP* problem for a given amount of available measurements. The characterization of the uncertainty is a central piece of this formalization. In particular, the *OSP* problem is addressed from the perspective of minimizing the remaining field uncertainty and sequential algorithms are proposed to solve it.

The use of *information theoretic (IT)* concepts such as *conditional entropy* has been studied to characterize the uncertainty related to a geological model conditioned to the acquisition of data (well logs), and its application in a preferential sampling strategy oriented to improve geostatistical inference at low acquisition rates. The conjecture has been that locations based on *IT-OSP* are distributed on transition zones of categorical fields, achieving better performance in tasks of image recovery than standard classical non-adaptive sensing schemes.

In the experimental side, a regularized greedy sequential algorithm is proposed to approximate the proposed *IT-OSP* sampling to show this principle. The proposed approach provides realizations based on multiple point simulations with reduced variability for geological categorical facies models in the critical low sampling regime.

Finally, the performances of different inference processes under the proposed sampling strategies are evaluated in some practical realistic scenarios for tasks related with grade control in short term planning.

"The scientific theorist is not to be envied. For Nature, or more precisely experiment, is an exorable and not very friendly judge of his work. It never says 'yes' to a theory. In the most favorable cases it says 'Maybe', and in the great majority of cases simply 'No'. If an experiment agrees with a theory it means for the latter 'Maybe', and if it does not agree it means 'No'. Probably every theory will some day experience its 'No' - most theories, soon after conception."

*Entry into memory book for Professor Kammerling-Onnes, November 11, 1922.
Quoted in Dukas and Hoffmann, Albert Einstein, the Human Side, p.18.*

*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.*

Agradecimientos

Inscrutable mysteries underlying the horizon of our journey. Are often shady the paths that we must go. Many of our experiences stained our memories of greyish colors with ink that can not be changed. Still, certain tones, certain sounds, remain rooted in the foundations of memory working as anchors that prevent completely and eternally sink us. Some faces, some voices, some smiles, some advice, or an occasional correction are always pleasant company in our memories and encouragement when it comes to achieving the goals, to undertake new challenges and face major changes.

Thank God, eternal force that has design and order to this universe, so wonderful and mysterious, which again and again provides us with concerns and mysteries able to capture, inundate and complement the restless mind with which we have been provided.

Gracias a Gloria Calabrán Contreras, amante, esposa y conviviente civil. Por compartir nuestro día a día. Por nuestras sobrinas. Por la paciencia y apoyo. Por la vida en conjunto, en la que día a día trabajamos como equipo. Thanks grandma and aunt for host and support when the rest of the universe looked in another direction, thanks a lot.

I strongly thank my advisors, Dr Jorge Silva and Dr Julian Ortiz, and to Dr Milan Derpich, who have allowed me to develop and formulate innovative and relevant concepts. Thanks for a collaborative thesis in areas of great importance to me, and for your invaluable support and for the various tasks entrusted. Thanks Thomas Peet and Ivan Castro and Jorge Toledo by helping and sharing with me along the formalization of subjects for this work.

Special thanks to The Coquetos and its support without which this dissertation would not have been possible. Muchas gracias Srta Andrea, por sus animadas canciones sin fin. Gracias por su chaleco verde que mantenemos en IDS y que nos acompaña a nuestras salidas, y que por sobre todo nos inspira a mejorar cada día y a rectificar continuamente nuestros infinitos errores. Un abrazo al Gran Sebita PRESO (&GOD), ayudante incansable de las futuras generaciones, amante de la enseñaza y la educación temprana. Alma noble preocupada por estudiantas, mechonas, y exmechonas de tantas generaciones. Como olvidar sus regalos misteriosos y las hordas de estudiantas esperando sus sabios consejos y su coqueta sonrisa y mirada. I hope one day to be able to see you two again, to embrace you both, to kiss both of you, and to hear her song "Abuelito dime tú".

Finally, to you, that for one reason or another uses this thesis, ¡Hasta Siempre!

Felipe Andrés Santibáñez Leal

Contents

1	Introduction	1
1.1	Context	1
1.1.1	Inverse Problem and Sensing	2
1.1.2	Inverse Problems in the Context of Sampling	2
1.1.3	Inverse Problems in Geosciences: A Sampling Problem	3
1.1.4	Geosciences and Uncertainty	4
1.2	Problem Statement	5
1.2.1	Channelized Structures	6
1.2.2	The Problem	6
1.3	Hypotheses	7
1.4	Objectives	7
1.4.1	Main Objective	7
1.4.2	Specific Objectives	7
1.5	Contribution of this Thesis	8
1.5.1	Summary of the results	9
1.6	Thesis Organization	10
2	General Background and Related Topics	11
2.1	Geostatistical Analysis and Inference	11
2.1.1	Two-Point Statistics	12
2.1.2	Object-based Simulation	13
2.1.3	Multi-Point Simulations <i>MPS</i>	13
2.1.4	Comparison of the Inference Techniques	14
2.2	Methods used for Sensing Design in <i>Geostatistics</i>	14
2.2.1	Sensor Placement Strategies	15
2.2.2	Another Sampling Techniques	15
3	Sampling Strategies for Uncertainty Reduction in Categorical Random Fields: Formulation, Mathematical Analysis and Application to Multiple-Point Simulations	17
3.1	Introduction	18
3.2	Related Works on Geological Uncertainty and Sampling Design	20
3.3	Preliminaries: Notation and Basic Concepts	22
3.3.1	Information Measures	22
3.3.2	Entropy as an Indicator of Simulation Complexity	23

3.3.3	A Basic Sensing Decision Problem	23
3.4	Sampling Strategies for Uncertainty Reduction	24
3.4.1	Problem Setting	24
3.4.2	The Equivalent Maximum Information Strategy	25
3.4.3	The Iterative Sequential Rule	26
3.5	The Adaptive Sensing Problem	29
3.5.1	Pseudo code to implement the Adaptive Rule	30
3.5.2	Analysis of the Information Gain	31
3.6	Case Study: A Binary Markov Chain	34
3.6.1	The Sequential Maximum Information Strategy (SMIS)	34
3.6.2	Empirical Analysis	35
3.6.3	The Adaptive Strategy AMIS	37
3.7	Application to Multiple-Point Simulations	38
3.7.1	Proposed Preferential Sampling Strategy for MPS	40
3.7.2	Practical Considerations for the Estimation of $H(X_{i,j} X_{S_{k-1}} = x_{S_{k-1}})$	41
3.7.3	Pseudo-code of the RAMIS solution	42
3.7.4	Experimental Setting and Validation	43
3.7.5	Selecting the Regularization Parameter α	44
3.7.6	Comparison with non preferential sampling methods	47
3.8	Discussion and Final Remarks	53
4	Improving Short-Term Planning and Grade Control with an Entropy- based Adaptive Sampling Strategy	54
4.1	Entropy-based Adaptive Sampling Strategy	56
4.1.1	Maximum Entropy Sampling	56
4.2	Optimal Sampling for Grade Control	59
4.2.1	Proposed Methodology for Sampling	59
4.2.2	Classification Process	60
4.3	Case Studies and Experiments	61
4.3.1	Database Description	61
4.3.2	Construction of Validation Block Model	63
4.3.3	Experimental Results: Adaptive Sampling Strategy	65
4.3.4	Performance Assessment	68
4.4	Chapter Conclusions	70
5	Conclusions	72
5.1	Future Work	73
A	Complementary Material for Chapter 3	74
A.1	App. I: Entropy as an Indicator of Simulation Complexity	74
A.1.1	The Operational Complexity Indicator for Simulation	74
A.1.2	Derivation of The Representation Result in (A.3)	75
A.2	App. II: Proof of Proposition 1	77
A.3	App. III: Optimality of the Sequential Rule for Fields with no Inter-Pixel Dependency	78

A.4	App. IV: Analysis of the Adaptive Sampling Scheme for the Markov Chain Case	79
A.5	App. V: Analysis of the Partial Update of Conditional Entropies	80
B	Complementary Material for Chapter 4	83
B.1	App. I: Pseudocode	83
B.2	App. II: Detailed Database	85
B.2.1	Case Study 2	85
B.2.2	Case Study 3	86
B.3	App. III: Additional Experimental Results	87
B.4	App. IV: Economic Evaluation	94
C	Complementary Information and Development	96
C.1	Details on <i>MPS</i>	96
C.1.1	<i>MPS</i> and the role of Training Images	96
C.1.2	Statistics from <i>MPS</i> Process	97
C.2	Details of additional methods for Signal Recovery	97
C.2.1	Methods used in Sparse Signal Recovery	97
C.3	Resolvability Capacity of the random field X	102
C.3.1	Resolvability Capacity of the Iterated Principle	102
C.3.2	RAMIS and Resolvability Capacity Properties	103
C.3.3	Stopping Rule via \mathcal{C}_K Evolution	104
C.4	On Regionalized variables with spatial dependence	106
C.4.1	<i>MRF</i> models and <i>Clique</i> structure estimation	106
C.4.2	<i>Clique</i> estimation and Mutual Information	106
C.4.3	Multi-point and Two-Point Mutual Information relation	107
C.5	Markovian 1D Scenario	109
C.5.1	<i>1-D</i> Markovian Spatial model	109
C.6	Formulation and Implementation of Noisy Sparse Promoting Solvers	116
C.6.1	<i>NCS</i> and <i>Whitening</i> process	116
C.6.2	On Noisy Compressive Sensing Performance	118
	Bibliography	125
	Nomenclature	135
	List of terms	136

List of Tables

1.1	Forward and Inverse problems in engineering and science	1
3.1	Global Performance Improvement after sampling level = 1.25%	47
4.1	Summary statistics.	62
4.2	Case study coordinates.	62
4.3	Summary Parameters SNESIM.	63
4.4	Performance Error Summary for Case Study 1.	68
4.5	Performance Error Summary for Case Study 2.	68
4.6	Performance Error Summary for Case Study 3.	68
4.7	Economic Profit Estimation for Case Study 1	69
4.8	Economic Profit Estimation for Case Study 2	69
4.9	Economic Profit Estimation for Case Study 3	70
B.1	Variables Considered in the Profit Analysis.	94
C.1	Summary of classical greedy approaches	99

List of Figures

1.1	Example of a 3-D geological reservoir.	3
1.2	Sampling Devices in Mining.	4
1.3	Categorical permeability channel.	6
1.4	Basic inverse problem scheme.	6
2.1	Example of a classical variographic analysis.	12
2.2	Example of a simulation based system.	14
2.3	Sensing design example and its effect on reconstruction of the field.	15
3.1	Diagram of the inputs and the required statistical model	28
3.2	Flow diagram to determine the sampling rule $\tilde{f}_k^*(\cdot)$	29
3.3	Diagram to solve the adaptive rule in Eq. (3.34)	32
3.4	Flow diagram to determine the sampling rule in Eq. 3.37	33
3.5	Conditional entropies for a binary Markov chain.	36
3.6	Uncertainty reduction for 20 realizations of the Markov chain.	38
3.7	Distribution of the conditional entropy of non-sensed locations.	39
3.8	Proposed training images.	42
3.9	Estimated maps for the model SC_1 at $k = 100$	45
3.10	Performance of non-sensed positions under RAMIS as a function of α	46
3.11	Resolvability of non-sensed positions using RAMIS approach.	48
3.12	Example of the masks used to define transitions in channelized images.	48
3.13	Experimental performance for model SC_1	49
3.14	Remaining Entropy maps for model SC_1	50
3.15	Experimental performance for model MC_1	51
3.16	Experimental performance for the model MC_2	52
3.17	Remaining Entropy maps from 600 realizations for model MC_1	52
3.18	Remaining Entropy maps from 600 realizations for model MC_2	53
4.1	Schematic diagram of the sampling rule (4.10).	61
4.2	Grade distributions and statistics for the selected databases.	63
4.3	Drill-hole samples data for case study 1. From left to right: Benches 1-6. Colormap denotes the grade of Cu	64
4.4	Blast-hole data for case study 1. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.	64
4.5	Ground truth images. Case study 1.	64

4.6	Samples for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using <i>Cut-Off</i> grade 1.012%. Colormap: Describe batch of samples in the order of the performed sampling.	65
4.7	Estimated grade for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using <i>Cut-Off</i> grade 1.012%. Colormap: the same as in Fig. 4.3.	66
4.8	Estimated grade control for Case Study 1. From left to right: Benches 2-6. From top to bottom: Ground truth, structured sampling, adaptive sampling using <i>Cut-Off</i> grade 1.012%.	67
4.9	Confusion Matrix for Case Study 1. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 1.012%.	67
A.1	Remaining conditional entropy for symmetric transition matrix ($\beta = 0.2$)	80
A.2	Estimation error for symmetric transition matrix with $\beta = 0.1$	81
B.1	Drill-hole samples data for case study 2. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.	85
B.2	Blast-hole data for case study 2. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.	85
B.3	Ground truth estimated from drill-holes and blast-holes for case study 2. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.	85
B.4	Drill-hole samples data for case study 3. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.	86
B.5	Blast-hole data for case study 3. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.	86
B.6	Ground truth estimated from drill-holes and blast-holes for case study 3. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.	86
B.7	Samples for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using <i>Cut-Off</i> grade 1.241%.	87
B.8	Estimated grade for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using <i>Cut-Off</i> grade 1.241%.	87
B.9	Estimated grade control for Case Study 1. From left to right: Benches 2-6. From top to bottom: Ground truth, structured sampling, adaptive sampling using <i>Cut-Off</i> grade 1.241%.	88
B.10	Confusion Matrix for Case Study 1. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 1.241%.	88
B.11	Samples for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using <i>Cut-Off</i> grade 1.518%.	89
B.12	Estimated grade for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using <i>Cut-Off</i> grade 1.518%.	89

B.13	Estimated grade control for Case Study 1. From left to right: Benches 2-6. From top to bottom: Ground truth, structured sampling, adaptive sampling using <i>Cut-Off</i> grade 1.518%.	90
B.14	Confusion Matrix for Case Study 1. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 1.518%.	90
B.15	Confusion Matrix for Case Study 2. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 0.22%.	91
B.16	Confusion Matrix for Case Study 2. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 0.445%.	91
B.17	Confusion Matrix for Case Study 2. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 0.692%.	92
B.18	Confusion Matrix for Case Study 3. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 0.115%.	92
B.19	Confusion Matrix for Case Study 3. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 0.273%.	93
B.20	Confusion Matrix for Case Study 3. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using <i>Cut-Off</i> grade 0.486%.	93
C.1	Synthetic image by <i>MRF</i> modeling	104
C.2	1-D regionalized variable.	109
C.3	Measured Variables in separated past and future subsets.	111
C.4	Statistics from simulations.	119
C.5	Performance Analysis of Naive Approach of <i>NCS</i> by relaxing restrictions	120
C.6	Examples of outcomes for <i>NCS</i> without considering spatial dependence.	121
C.7	Performance Analysis of Naive Approach of <i>NCS</i> adding <i>TV</i>	122
C.8	Example of <i>NCS</i> reconstruction under independence: SC1	122
C.9	Example of <i>NCS</i> reconstruction under independence: MC1	123
C.10	Example 1 of <i>NCS</i> reconstruction under independence: MC2	123
C.11	Example 2 of <i>NCS</i> reconstruction under independence: MC2	123
C.12	Example 1 of <i>NCS</i> with full covariance estimation: MC2.	124
C.13	Example 2 of <i>NCS</i> with full covariance estimation: MC2.	124

Chapter 1

Introduction

1.1 Context

In complex real scenarios the task of recovering an underlying phenomenon or system (for example, a variable from an industrial process, or a model or a mathematical object of an environment to be studied) from measurements could be extremely hard and surrounded by numerous sources of uncertainty (Boucheron, Stéphane et al, 2005; Chiles and Delfiner, 2012; Cover and Thomas, 2006; Gray and Davisson, 2004; Kay, 1993; Kochenderfer et al, 2015; Moon and Stiriling, 2007; Scheidt and Caers, 2009; Vapnik, 1999). In general, this task could be formulated as an inverse problem and relies on the interplay between the acquired information of the system (termed data or measurements), and the assumptions made about the underlying model, variable or phenomenon to be estimated (typically summarized in terms of a set of parameters). Some general categories for this problem are summarized in Table 1.1. Typically, the inverse problem reduces to finding the best or simplest explanation from the data, describing an attempt to estimate a model or phenomenon coherent with the available evidence (Santamarina and Fratta, 2005). In this context, prior knowledge about the nature of the problem implies describing in mathematical terms the information available about the problem that is complementary to data.

Table 1.1: Forward and Inverse problems in engineering and science

	Forward Problems		Inverse Problems	
	System Design	Convolution	System identification	Deconvolution
Input:	Known	Known	Known	Unknown
System:	To be designed	Known	Unknown	Known
Output:	Predefined	Unknown	Known	Known

1.1.1 Inverse Problem and Sensing

In an inverse problem, the general relationship between the distribution of the set of parameters to be estimated (denoted by \mathbf{X} and related with the target hidden variable) and the observations or measurements, denoted by \mathbf{Y} , is often described by a complex and non-linear forward relationship given by the function $g(\cdot)$ in Eq. (1.1). Therefore the data acquisition process (a *sensing* process) is described by:

$$\mathbf{Y} = g(\mathbf{X}). \quad (1.1)$$

Some level of uncertainty can be added in the relationship (1.1) by incorporating a noise component, obtaining a more realistic observation model that offers a probabilistic mapping between \mathbf{Y} and \mathbf{X} .

$$\mathbf{Y} = g(\mathbf{X}) + \nu. \quad (1.2)$$

Then the inverse problem can be formulated as follows: given Y , consistent with a forward model $g(\mathbf{X})$, the objective is to estimate \mathbf{X} with some criterion. As \mathbf{X} is unknown, an objective function that measures the discrepancies or error between Y and $g(\mathbf{X})$ is required. The standard metric is the square error. In addition, the problem could be ill-posed meaning that many possible sets of solutions are consistent with Y and the observation model.

In this context, a classical objective function to implement the inverse problem has the following form:

$$G(\mathbf{Y}, \mathbf{X}, g(\cdot)) = \|\mathbf{Y} - g(\mathbf{X})\|_p, \quad (1.3)$$

where $\|\cdot\|_p$ denotes the p norm. In addition, when the problem is ill-posed some form of regularization is added to Eq. (1.3) (Davis et al, 1995; Magnant, 2011).

1.1.2 Inverse Problems in the Context of Sampling

In a wide range of applications, \mathbf{X} belongs to a high dimensional space, which makes unfeasible or at least impractical its full observation. Thus, considering \mathbf{X} as a signal residing in a high dimensional space \mathbb{R}^N , only a finite number of measurements m is available with $m \ll N$, $\mathbf{Y} \in \mathbb{R}^m$, obtaining the following sampling model:

$$\mathbf{Y} = A(\mathbf{X}) + \eta, \quad (1.4)$$

where the function $A(\cdot)$ represents the sampling scheme and η is the noise associated to the measurements.

Traditional linear sampling signal processing theory (Proakis and Manolakis, 1996)[section 1.4.2] states that data acquisition systems require to sample signals at a rate exceeding twice the highest spatial/temporal frequency for the purpose of losslessly recovering a band limited signal. This is the principle behind the majority of imaging acquisition and audio recorders. However, in many practical problems only $m \ll N$ measurements are accessible to recover \mathbf{X} (i.e. solve the ill-posed problem) from Eq. (1.4), and consequently an error in the recovery is introduced.

1.1.3 Inverse Problems in Geosciences: A Sampling Problem

In the context of many relevant inverse problems in *Geosciences*, the relationship expressed by a model like Eq. (1.4) needs to be stipulated from a physical model, or estimated from empirical evidence and a statistical model that connects the observed (sampled) data \mathbf{Y} from a set of parameters representing \mathbf{X} .

An important task and the main focus of this thesis is the problem of reservoir characterization (Kitanidis, 1997; Oliver et al, 2008; Onwunalu and Durlofsky, 2009; Strebelle and Zhang, 2004). In the characterization of a reservoir, as the one shown in Fig. 1.1, several variables are relevant to describe a physical phenomenon including discrete ones such as fluid filling indicators, rock or sediments types, and continuous ones such as porosity and permeability (Kitanidis, 1997; Oliver et al, 2008; Strebelle and Zhang, 2004). One of the main challenges lies on the fact that no direct observations (or just a reduced amount of these observations) are available leading to the use of indirect correlated data for the inference of \mathbf{X} .



Figure 1.1: Example of a 3-D geological reservoir. By Minnesota Control Pollution Agency, <https://www.pca.state.mn.us/karst-outreach>. Accessed 01 May 2019.

In *Geosciences*, an accurate description of the spatial distribution of a subsurface model is essential for reservoir characterization, which plays a key role for exploration and production in the mining industry (Bangerth et al, 2005, 2006; Güyagüler Baris, 2002; Krause et al, 2006; Oliver et al, 2008; Onwunalu and Durlofsky, 2009). In the context of the model in Eq. (1.4), the limited access to measurements makes the inference of \mathbf{X} a very challenging problem. In particular, the aforementioned sampling problem implies that geophysical inverse problems are typically undetermined. Therefore, characterizing a reservoir as the

one shown in Fig. 1.1 up to some level of accuracy requires the use of several sources of information, such as: *wells* (direct samples), production data, and seismic data (Güyagüler Baris, 2002; Krause et al, 2008a; Olea, 1984; Oliver et al, 2008). *Seismic data* is usually available at large scale (low spatial resolution) and it is provided for the entire reservoir. However, this data is sampled on a coarse grid with several sources of uncertainties and noise. On the other hand, *Well observations* consist on well logs taken from a process illustrated in Fig. 1.2. *Wells* are sampled on a fine grid in a specific path of interest providing excellent spatial resolution. The uncertainty associated with these measurements is significantly smaller than seismic data. However, *well data* is sparse and only available in very restricted areas (provided by the existing wells) because it is a very expensive process.

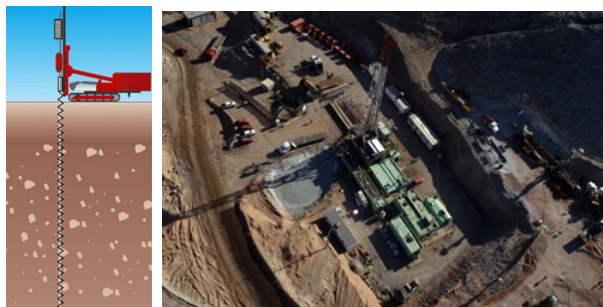


Figure 1.2: Sampling Devices in Mining.

Left: Example of a well sampling scheme. Right: Example of an actual well sampling system. Left image by GCZ ingenieros S.A.C. 2016, <http://www.gczingenieros.com/productos.php?id=18>, Accessed 01 May 2019. Right image by RTV.es, <http://www.rtv.es/fotogalerias/odisea-33-mineros-atrapados-chile/56175/maquina-del-plan-alcanza-519-metros-rescate-33-mineros-chile/29/>, Accessed 01 May 2019

In the case of *production data*, the information is acquired along the production process in *production wells*. This kind of data takes into account global factors related with features of reservoir that makes a great difference with the other two sensing modalities (Abellan and Noetinger, 2010; Handwerger et al, 2016; Man and Jarzyna, 2013; Oliver et al, 2008; Raef et al, 2015; Strebelle and Zhang, 2004). This information, also known as *historical data*, considers numerous scenarios which produces a large volume of information about the system. This information has been used very successfully to implement predictive based inference such as the history matching approaches (Kitanidis, 1997; Oliver et al, 2008; Scheidt and Caers, 2009). It is important to mention that although it is a very relevant and interesting topic, production data has not been used or considered for the purpose of this thesis.

1.1.4 Geosciences and Uncertainty

As it was mentioned before, the information sources about the reservoir characterization contain uncertainties (Güyagüler Baris, 2002; Onwunalu and Durlofsky, 2009; Wellmann, 2013). Therefore, several techniques have been developed in order to estimate, quantify, and represent these uncertainties (Bangerth et al, 2006; Krause et al, 2008a). In general,

the access to noiseless observations of the subsurface structures is not possible because direct measurements are limited in number and these are unevenly distributed. Based on the above, the geological characterization is performed by using several indirect information sources (Kitanidis, 1997; Oliver et al, 2008). In this context, a stochastic modeling of the problem is essential to capture the uncertainties in the inference process (Bangerth et al, 2006).

Reservoir properties at various grid locations (pixels on a discrete two or three dimensional representation) are largely unknown, hence each property of interest at every grid block (or pixel) is modeled as a random variable whose variability is described by a probability measure (Blackwell, 1998; Brus and Heuvelink, 2007; Gao et al, 1996; Lloyd and Atkinson, 1998). The reservoir characterization relies not only on a reduced portion of available data but also on their spatial disposition. This rises the importance of sensing placement as a task that is intrinsic in the formulation of the inverse problem in Eq. (1.4). This task of optimal sensing design is the main focus of this thesis.

1.2 Problem Statement

In this thesis, the focus is on the characterization of subsurface structures from spatial observations and the role that *sampling* and *inference strategies* could jointly have in this task.

For *Geosciences*, the global outcome corresponds to the characterization of an unknown complex reservoir field by the description of its individual components (such as high number of structures, facies¹ and properties in the subsurface). The focus of this work is the recovery of a single subsurface property described by a *2-D* regularized variable sampled in the spatial domain (see Fig. 1.4). The characterization of these kinds of structures is relevant for mineral prospecting, mine planning, and production stages where the number of available observations is severely restricted by technical and economical factors. Based on the above, the integration of adaptive sensing schemes and an appropriate inference methodology provides a promising solution to improve classical geostatistical inference that are based on non-adaptive sampling strategies (Güyagüler Baris, 2002; Krause, 2008; Krause et al, 2008a, 2009).

Beyond the prospecting context described above, both exploitation by mineral blasting and short-term mine planning could also benefit from the use of an adaptive sensing strategy for inference. While blast hole drilling systems has been focused on efficient drilling instead of high precision sampling, short term mine planning uses a medium term sampling strategy to choose mining units based on estimated distribution of ore and waste. Units classified as ore will be sent to the plant while waste units will be sent to the waste dump. Mistakes in this classification task have a significant impact on economic outcome of the productive process, and an adaptive sensing strategy could improve the inference and consequently the economic impact of a mining project.

¹rock structures recognized by its composition or fossil content and mapped by these characteristics.

1.2.1 Channelized Structures

In the first part of this thesis, a classical geosciences scenario related with subsurface channels systems is explored. As previously shown in Fig. 1.1, several continuous or discrete variables could be related with channelized structures, where characterizing the extent and location of channels is relevant to infer petrophysical properties of the rock structure (porosity, permeability, etc.). In addition, a statistical model is used as part of the forward model describing the channelized structures, where a non-stationary assumption is needed to reproduce channel-like features (Oliver et al, 2008; Remy et al, 2009). An example of a realization of this model is presented in (Fig. 1.3).



Figure 1.3: Categorical permeability channel.
Example of a 2-D representation of a categorical channelized permeability field.

1.2.2 The Problem

Formally, the image model is a numerical representation of the spatial distribution of a subsurface attribute, such as thickness, permeability, porosity, or flow rate (Kitanidis, 1997; Lam, 1983; Matheron, 1971). The general process of image reconstruction or characterization (i.e. inference of the spatial distribution) is described in Fig. 1.4, where given a reduced number of observations the main goal is to infer the underlying image or some significant feature such as first and second order statistics, connectivity, or transport metrics of the image.

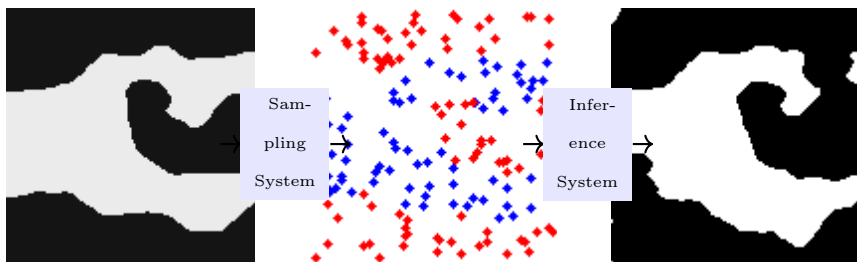


Figure 1.4: Basic inverse problem scheme.
Example of a basic scheme of inverse problem related with 2-D channelized structures characterization.

The inverse problem of characterizing a field based on few measurements can be assessed from two angles: i) from the point of view of the inference (i.e. given an appropriate

sensing scheme) the target is to infer some global and local properties of the field everywhere from the samples; ii) from the point of view of experimental design or sensing design, the optimal sensing schemes optimizing $m \ll N$ observations is the problem of finding the best positions for the observations in order to maximize the information to recover the signal \mathbf{X} . This problem is usually termed as *Optimal Sensor Placement (OSP)* (Krause, 2008; Krause et al, 2011; Olea, 1984).

1.3 Hypotheses

The main hypotheses of this work are:

1. 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.
2. Adaptive sensing schemes can be integrated in the inference to improve the state-of-the-art of geological field characterization.
3. 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.

1.4 Objectives

1.4.1 Main Objective

The main objective for this research is to enhance the reconstruction of images describing $2-D$ categorical regionalized variables by the use of new *sensing* strategies that takes into account uncertainty reduction as a criterion and the use of previously sampled data and side information from a statistical model of the field. The focus of this research is on regionalized variables with several spatial dependence assumptions by taking advantage of its spatial structure and other sources of expert knowledge of the media of interest.

1.4.2 Specific Objectives

The specific objectives addressed in this work are:

- 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.

1.5 Contribution of this Thesis

In this thesis, a new information driven formulation of a sampling strategy is proposed by formalizing some of the preliminary ideas introduced by Wellmann (2013). In particular, a sampling framework that integrates both sampling data in a sequential scheme and the statistical information obtained from training images in the context of *MPS* is proposed. The use of maximum entropy as a criterion for preferential sampling is proposed in this thesis as a concrete alternative to classical non-adaptive sampling schemes.

On the specifics, a new adaptive sequential empirical maximum entropy sampling (*AdSEMES*)² approach is formulated and implemented by integrating *information theoretic* concepts. This sampling strategy can be seen as an adaptive way to locate the samples on the positions that maximizes discrimination about the transition zones of facies that determines the underlying phenomenon. In particular, three maximum entropy sampling frameworks have been formalized and implemented within the context of regionalized categorical variables with spatial dependencies.

In specific, the contributions presented in this work answer the following questions.

- Given K available measurements, what is the *best* disposition of these measurements from an information perspective (in the sense of uncertainty reduction)?
- Given an image model, is there a sampling regime where optimal sampling can outperform classical sampling approaches?
- How the previous gain is determined by the complexity of the statistical model of the field?
- What is the best inference methodology for the proposed optimal sensing strategy?
- If the inference methodology is based on *MPS*, can entropy and mutual information be good predictors of its performances?

²Or regularized adaptive maximum information sampling (*RAMIS*) strategy.

1.5.1 Summary of the results

The main contribution of this research is the formalization of the task of optimal sampling for the statistical simulation of a discrete random field addressed from the perspective of minimizing the posterior uncertainty of non-sensed positions given the information of the sensed positions. In particular, information theoretic measures are adopted to formalize the problem of optimal sampling design for field characterization, where concepts like information of the measurements, average posterior uncertainty, and the resolvability of the field are introduced. The use of the entropy and related information measures are justified by connecting the task of simulation with a source coding problem, where it is well known that entropy offers a fundamental performance limit. On the application, a one-dimensional Markov chain model is explored where the statistics of the random object are known, and then the more relevant case of multiple-point simulations of channelized facies fields is studied, adopting in this case a training image to infer the statistics of a non-parametric model. On both contexts, the superiority of information-driven sampling strategies is proved in different settings and conditions, with respect to random or regular sampling. The details of this contribution are presented in Chapter 3, and a published article (**Santibañez † F** et al, 2019).

The second main contribution of this thesis is a method to select sampling locations in an advanced drilling grid for short-term planning and grade control in order to improve the correct assessment (ore-waste discrimination) of blocks. The sampling strategy is based on a regularized maximization of the conditional entropy of the field, an objective function that formally combines global characterization of the field with the principle of maximizing information extraction for ore-waste discrimination. This sampling strategy is applied to three real cases, where dense blasthole data is available for validation from several benches. Results show relevant and systematic improvement with respect to the standard regular grid strategy, where for deeper benches the gains in image inference are more prominent. The details of this contribution are presented in Chapter 4 of this thesis. These results are part of a scientific article (**Santibañez-Leal † F** et al, 2019).

Publications on ISI Journals and International Conferences

- Journal article 2019a: **Santibañez † F**, Silva JF, Ortiz JM (2019) Sampling strategies for uncertainty reduction in categorical random fields: Formulation, mathematical analysis and application to multiple-point simulations. *Mathematical Geosciences* 51:579–624. Published.
- Journal article 2019b: **Santibañez-Leal † F**, Silva JF, Ortiz JM (2019) Improving short term planning and grade control with an entropy-based adaptive sampling strategy. *Natural Resources Research* online first:1–35. Submitted, Under revision.
- Journal article 2019: Calderon* H, **Santibañez* † F**, Silva JF, Ortiz JM, Egaña A (2019) Geological facies recovery based on weighted ℓ_1 -regularization. *Mathematical Geosciences* online first:1–46. Accepted.
- International Conference Paper 2016: Calderón H, Santibanez F, Silva JF, Ortiz JM, Egaña A (2016) Channelized facies recovery based on weighted compressed

sensing. In: 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Rio de Janeiro, Brazil, July 10-13, 2016, pp 1–5, doi: 10.1109/SAM.2016.7569627, URL <https://doi.org/10.1109/SAM.2016.7569627>. Published. Oral presentation performed by Santibáñez. Rio de Janeiro, Brasil. Jul. 2016.

1.6 Thesis Organization

The rest of this thesis is organized in two main chapters that present the main contribution of this work.

Chapter 2 summarized the general background in *Geostatistics* related with the research performed in this work.

Chapter 3 introduces the formalization of the optimal sampling decision problem and the proposed regularized adaptive sampling strategy. The relation between sampling strategies and uncertainty reduction is described in order to formulate the proposed adaptive sampling methodology. Both analytic and experimental studies are conducted on different scenarios. Finally some practical considerations are addressed and results are presented that demonstrated the advantages of the propose method.

Chapter 4 presents a new methodology for optimal sampling in the context of ore-waste discrimination. The problem of optimal sampling is formalized in this binary decision context and a novel methodology is proposed. The objective considers both the image recovery and the economic impact of the project.

Chapter 5 concludes with a summary of this thesis work, and offers directions and discussion for future research. Finally, the appendices contain complementary material that support the presentation of this work.

Chapter 2

General Background and Related Topics

In *Geostatistics* the community has developed several statistics tools in order to achieve good estimations for describing structures with spatial dependence such as channelized fields. In this section the *state-of-art* of multi-point simulations (Sec. 2.1), and sensing design (Sec. 2.2) is summarized.

2.1 Geostatistical Analysis and Inference

In recent decades, geologists have achieved realistic representations of the internal structure of reservoirs considering complex and heterogeneous geological environments through the use of *Geostatistics* (Arpat and Caers, 2007b; Bittencourt and Horne, 1997; Holden et al, 1998; Journel and Kyriakidis, 2004; Man and Jarzyna, 2013; Strebelle and Zhang, 2004). *Geostatistics* deal with spatially correlated data such as geological facies, reservoir thickness, porosity, and permeability (Calderón et al, 2016; Calderon* et al, 2019; Oliver et al, 2008; Raef et al, 2015). *Geostatistics* tools allow to evaluate potential exploration of a reservoir and its production, where one of the main issues is to determine relevant samples location.

More specifically, the term *Geostatistics* refers to a branch of spatial statistics that is concerned with the analysis of an unobserved spatial phenomenon $X = \{X_{(u,v)} : (u, v) \in D \subset \mathbb{R}^2\}$ (for the 2-D case), where D denotes a geographical region of interest (see Fig. 1.1). When the spatial coordinates are discretized the subset D is comprised by only N positions allowing the representation of the field as the collection $X = \{X_i : i \in \{1, \dots, N\}\}$. Therefore, *Geostatistics* deal with stochastic processes defined in a region D , with $D \subset \mathbb{R}^d$ and considering $d = 1, 2$, or 3 . For continuous valued stochastic fields, a classical assumption is that the field X is modeled by a stationary and isotropic *Gaussian process* (*GP*), with zero mean, constant variance σ^2 and autocorrelation function $\rho(X_i, X_{i+h}; \phi) = \mathbb{E}\{X_i \cdot X_{i+h}\}$ given by $\rho(X_i, X_{i+h}; \phi) = \rho(\|h\|; \phi)$, $\forall \{i\} \in D$ (Abellan and

Noetinger, 2010; Blackwell, 1998; Gao et al, 1996; van Groenigen et al, 1999). Typically for inference, the *Regionalized Variable*(*RV*) X is known at M limited locations that is denoted by $X_f = \{X_{f(m)} : m = 1 : M\}$. The goal is to estimate from X_f at unsampled locations. At every location, the *RV* is interpreted as a random variable, the goal is to infer the conditional distribution of X at unsampled locations. More details about this inference process will be presented in Sec. C.1 of this document. For completeness, the following subsections provide a general overview of some of the tasks and techniques used for this inference process.

2.1.1 Two-Point Statistics

When large amount of measurements are available, it is a common practice to use interpolation techniques around the observed data (Kitanidis, 1997). One of these techniques is *Kriging* (Abellan and Noetinger, 2010; Blackwell, 1998; Gao et al, 1996; van Groenigen et al, 1999), *Kriging* is an interpolation technique that uses variogram of X (second order statistics) (Lam, 1983) as shown in Fig. 2.1. The variogram is a representation of the spatial correlation of a two or three dimensional model (Abellan and Noetinger, 2010). Classical *Kriging* approaches are based on *two-point* statistics (Kitanidis, 1997). The variographic analysis measures only facies continuities between two regionalized variables, failing in the effective representation of curvilinear or multiscale structures that requires the inference of joint correlations of facies at multiple variables positions. Thus, these techniques based on variograms reproduction tend to fail in the modeling of realistic geological facies that can not be represented by stationary Gaussian fields. In practice, these models are unable to properly represent long-range properties of subsurface fields, misrepresenting the real reservoir connectivity. This mismodelling issue translates in poor reservoir characterization for exploration and production tasks (Bangerth et al, 2005; Olea, 1984; Ortiz and Deutsch, 2004).

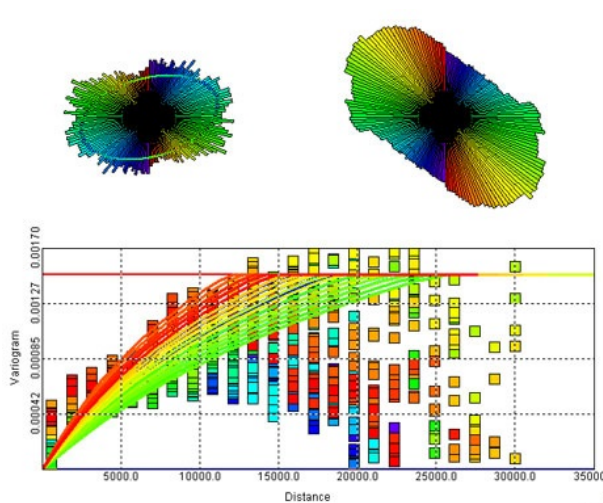


Figure 2.1: Example of a classical variographic analysis.

2.1.2 Object-based Simulation

Both geologists and engineers have a keen interest in local scale details describing reservoir heterogeneities. For this challenging task, stochastic simulations provide an appropriate tool with special emphasis in the regime when a small level of data is available (Kitanidis, 1997; Lam, 1983). An initial approach considered by the *Geostatistics* community for stochastic simulation has been based on object-based simulation (Chilés and Delfiner, 1999; Holden et al, 1998). This method simulates many spatial variables by the superposition of some predefined geometric patterns (e.g. discs, sinusoids, manifolds). Predefined geological shapes require to be manually selected by an expert. In contrast to the variographic analysis presented before, object-based methods provides realistic facies structures, but the selection and acquisition of an appropriate conditioning data is a critical limitation (Chilés and Delfiner, 1999; Holden et al, 1998).

2.1.3 Multi-Point Simulations *MPS*

In regimes of low data acquisition, the standard approach to the geostatistical inference involves the generation and analysis of multiple point simulations (*MPS*) (Huang et al, 2013; Ortiz and Deutsch, 2004; Scheidt and Caers, 2009). These are realizations of a spatial model conditioning on the available data (direct samples). This process is conducted by replicating patterns from a training image (Ortiz and Deutsch, 2004; Remy et al, 2009).

Geostatistical simulations provide a powerful and computationally efficient tool to reproduce more faithfully and realistically the spatial variability observed in a model when a small numbers of measurements are available (Ortiz and Deutsch, 2004; Remy et al, 2009). A simple illustration of this process is shown in Fig. 2.2. Another important dimension of this approach is that, Geostatistical simulations lead to many realizations of a reservoir model that can be conditioned to geologic, seismic and production data. These models (or realizations) provide appropriate representations of *geological heterogeneities* integrating various types of data at different scale and with different precisions. Finally, in any regime of data, expert knowledge is required to validate and to interpret the use of *Geostatistical* simulations (Bangerth et al, 2006; Oliver et al, 2008).

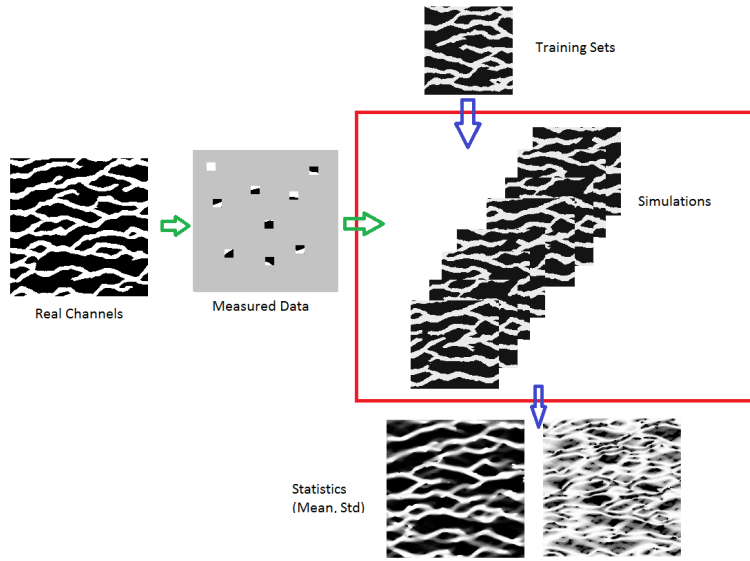


Figure 2.2: Example of a simulation based system.

2.1.4 Comparison of the Inference Techniques

Concerning deterministic *Kriging*, this technique provides solutions that are good close to conditioning data. Object-based approaches, on the other hand, provides acceptable solutions for the reproduction of geological shapes. However, none of these methods is good at both contexts (Bangerth et al, 2005; Lam, 1983; Olea, 1984; Ortiz and Deutsch, 2004). Classical *MPS* was developed to combine the strengths of the two previously stated approaches. *MPS* methodology recreates a realistic realization of the target field keeping the flexibility of pixel based simulation methods. At the same time, honoring well or seismic data is feasible at the initial stage of the simulated process. Furthermore, *MPS* realizations can reproduce complex and realistic geologic structures by the estimation of simultaneous statistics at multiple positions from the reference model (training image).

For completeness, more details about *MPS* can be found in the App. C.1

2.2 Methods used for Sensing Design in *Geostatistics*

The selection of the best informative observations for an inference task is a critical problem in statistical signal processing, with numerous applications in inference and decision: temperature and light monitoring (Davis et al, 1995; Krause, 2008), sensing contamination in a river (Christakos and Killam, 1993; Christodoulou et al, 2013; Gutjahr, 1991), mining exploration (Aspie and Barnes, 1990; McBratney and Webster, 1981; Ortiz and Magri, 2014), collaborative robotic networks design (Krause, 2008; Zidek et al, 2000), statistical experimental design (Magnant, 2011; McBratney et al, 1981; Olea, 1984; Vasát et al, 2010), among many others (Bangerth et al, 2006; Bittencourt and Horne, 1997; Guestrin

et al, 2005; Krause et al, 2006, 2008a; Xu et al, 2017). In general, this optimization problem is NP-hard, which is very critical at low acquisition rates.

2.2.1 Sensor Placement Strategies

In the context of facies recovery in *Geostatistics*, the focus has been on the scenario of pixel based measurements that model well logs. Here, the sensing design reduces to the optimal well placement (*OWP*) problem. In a nutshell, *OWP* finds the most proper locations to take measurements by drilling wells in the field of interest. The related more general problem of optimal sensor placement *OSP* is an active area of research in *communications* (Krause et al, 2009, 2011) and *machine learning* (Guestrin et al, 2005; Krause et al, 2008b). *OSP* (and *OWP* in particular) states the optimal (or near optimal) systematic way to take measurements in order to maximize the inference performance in some metric. For example, Fig. 2.3 illustrates the design in terms of two sensing approaches in a recovery based system.

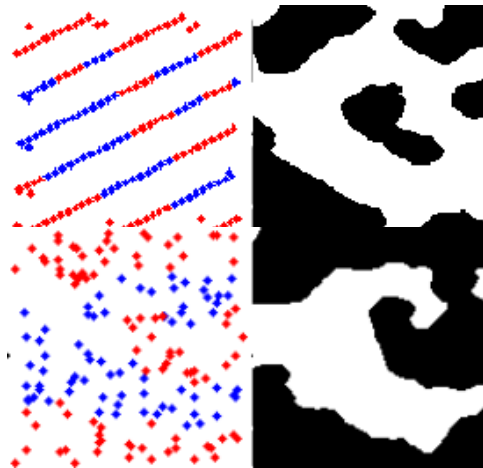


Figure 2.3: Sensing design example and its effect on reconstruction of the field.

Upper row: an arbitrary structured sensing scheme (left) and the achieved reconstruction by the measures at this scheme locations (right). Lower row: a near optimal sensing scheme (left) and the achieved reconstruction by the measures at this scheme locations (right). The real image corresponds to the fig. 1.3

2.2.2 Another Sampling Techniques

Some approaches in the literature have proposed preferential sampling schemes oriented to optimize productivity (production functionals) and economic factors (Bangerth et al, 2005, 2006; Onwunalu and Durlofsky, 2009). Several optimization methods have been proposed to achieve some of these optimal functionals such as adjoint-based gradient (Onwunalu and Durlofsky, 2009), simultaneous perturbation stochastic approximation (*SPSA*) (Bangerth et al, 2006), finite difference gradient (*FDG*) (Bangerth et al, 2006), very fast simulated annealing (*VFSA*) (Bangerth et al, 2006), binary genetic algorithm (*bGA*) (Onwunalu and Durlofsky, 2009), continuous or real-valued GA (*cGA*) (Onwunalu

and Durlofsky, 2009), and particle swarm optimization (*PSO*) (Onwunalu and Durlofsky, 2009).

While these economical and functional factors are important for specific applications, the idea of global uncertainty reduction is more fundamental and it could also have a positive impact on the improvement of economical and functional factors (Krause et al, 2006). In this context, Wellmann (Wellmann, 2013) proposed the use of *information theoretic* principles in the geostatistical analysis of a map making task. In this specific problem, a direct connection between *conditional entropy* and its applicability to the characterization of the uncertainty of regionalized variables has been shown. On the theoretical side, Wellmann (Wellmann, 2013) proposed that, in a categorical stochastic field, the regionalized variables with higher *conditional entropy* are located in the facies transition zones, implying that this information metric is relevant for detecting facies transitions.

A related topic in signal recovery in an under-sampled setting is presented in the App. C.2 for the case of sparse signal reconstruction based on linear measurements.

Chapter 3

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

The task of optimal sampling for the statistical simulation of a discrete random field is addressed from the perspective of minimizing the posterior uncertainty of non-sensed positions given the information of the sensed positions¹. In particular, information theoretic measures are adopted to formalize the problem of optimal sampling design for field characterization, where concepts like information of the measurements, average posterior uncertainty, and the resolvability of the field are introduced. The use of the entropy and related information measures are justified by connecting the task of simulation with a source coding problem, where it is well known that entropy offers a fundamental performance limit. On the application, a one-dimensional Markov chain model is explored where the statistics of the random object are known, and then the more relevant case of multiple-point simulations of channelized facies fields is studied, adopting in this case a training image to infer the statistics of a non-parametric model. On both contexts, the superiority of information-driven sampling strategies is proved in different settings and conditions, with respect to random or regular sampling.

¹The content of this chapter is based on the **Santibañez † F**, Silva JF, Ortiz JM (2019) Sampling strategies for uncertainty reduction in categorical random fields: Formulation, mathematical analysis and application to multiple-point simulations. *Mathematical Geosciences* 51:579–624. Published

3.1 Introduction

The recovery of an image from scarce measurements is at the core of many reconstruction problems in geosciences. In many contexts direct measurements are very expensive and consequently this inference task needs to be addressed in regimes where only scarce information from sensed data is available. In geostatistics, this lack of information from data has been addressed in the context of multiple-point statistics by the use of a prior statistical model estimated from a training image (Arpat and Caers, 2007a; Guardiano and Srivastava, 1993; Mariethoz and Caers, 2014; Ortiz and Deutsch, 2004; Strebelle, 2002; Wu et al, 2008). The idea is to use prior information to interpret the measurements and to provide realistic scenarios of the underlying field uncertainty. This prior information is represented by a probability model on the pixel space, the so-called multiple-point statistical model, that is used in conjunction with the available data to create independent and identically distributed (i.i.d.) realizations of the model conditioned on the measurements. In practice, MPS simulates non-sensed values in an inductive way by estimating the conditional probability of a non-sensed location given the values of the sensed and the previously simulated locations. This conditional probability is inferred from a training image. Therefore the recovery or characterization of non-sensed locations is obtained by a collection of conditional simulations where there is an intrinsic uncertainty that is captured by a set of feasible solutions (simulations). This stochastic recovery is justified from the fact that the focus is the regime of scarce measurements that translates in a significant posterior uncertainty after the measurements. It is evident that this high posterior uncertainty regime is not adequately represented by one solution, for instance the conditional mean attributed to the estimation problem (Gray and Davisson, 2004). Simulations, provide a more adequate way of representing the intrinsic uncertainty of the field and realistic features from the training image (Mariethoz and Caers, 2014).

In order to properly characterize the random field considered, samples must be representative. This is usually achieved by having a non-preferential sampling through a regular or quasi-regular grid (squared or rectangular) of sensed locations (Rossi and Deutsch, 2014) or a fully random scheme (Wellmer, 1998). However, in practice in the case of categorical fields, knowledge about the location of transitions between categories is key to predict the economic performance of the reservoir or mineral deposit. A classical approach considers a two-stage approach. First, a wide-space regular grid is sensed to delineate favorable areas, and then relevant areas are characterized with a denser, but still regular grid (Kennedy, 1990).

This work departs from this non-preferential principle and asks the question about the role of preferential sampling in MPS. In this context, the first objective is to evaluate the benefits of a preferential strategy and, secondly, to uncover the role of the training image in this task and how this model can be used to improve the decision of where to measure. The main conjecture is that there is a sampling regime where the role of the model is relevant to determine where to measure and that, consequently, an adaptive sampling strategy can improve the performance of MPS.

In order to formalize this idea, the problem of sampling design (or sensing placement)

is addressed from an information perspective (Cover and Thomas, 2006; MacKay, 2002), posting it as the problem of reducing the uncertainty of non-sensed positions given the information provided by the sensed positions. This sampling principle is inspired by the work of Wellmann (2013), where information theoretic measures are proposed as a way to quantify potential reduction in uncertainties given a new sample in the context of a general geological model. To formalize the notion of uncertainty and information in this simulation context, the role of the Shannon entropy is studied following the works of Wellmann (2013); Wellmann and Regenauer-Lieb (2012). In particular, it is shown that the entropy offers a way of expressing the complexity of the simulation task. For that, the simulation problem is interpreted as the problem of characterizing a typical set where its cardinality can be seen as an indicator of the complexity of this task (See App. A.1). Therefore, the sampling design can be posted as the problem of minimizing the entropy of the posterior distribution of non-sensed positions given the sensed data (Sects. 3.4 and 3.5). Interestingly, the optimal sampling (OS) problem in this framework is the problem of selecting the more informative positions, in the sense suggested in (Wellmann, 2013, Sec.4), that is the positions that after being measured provide (in average) the highest reduction in the uncertainty of the problem (Cover and Thomas, 2006).

On the application and validation of these ideas, a first exploration of a known scenario (one-dimensional Markov chain model) is studied where the entropy and related quantities can be computed with high precision. From this analysis, the benefits of preferential sampling are confirmed with respect to non-preferential solutions and, at the same time, insights are obtained into the way OS solutions are distributed in the pixel space. Interestingly, a link with quasi-regular sampling is presented as there is a case (non-adaptive) where this solution is optimal in the sense of reducing uncertainty under a Markov model assumption.

Moving on to the adoption of these sampling ideas in the task of multiple-point simulation (MPS), an information-driven sampling strategy is proposed that addresses some practical limitations that are not present in the idealized setting where the model is assumed to be known. The main difficulty in implementing the OS strategy in this context comes from the need to estimate the model statistics from a training image. This leads to two main issues: one associated with the statistical inference errors of pattern statistics and the other with the algorithmic complexity of inferring statistics as new sensed locations become available. To address them, a solution is proposed with a compound objective function that reflects a compromise between the global and local characterization of the field. The local component emphasizes the most informative non-sensed areas typically around the transition zones of the true field, while the global component enforces a more uniform sampling of the domain. This preferential scheme is evaluated for three channelized facies models and in a range of sampling rates. Results are compared with two non-preferential solutions. Results show concrete evidences of the benefits of preferential sampling both in the sense that the obtained simulations are closer (in average) to the true image (bias) and also in that simulations show less variability in the non-sensed positions measured in terms of the average conditional entropy (variance). As expected, the magnitude of the gain depends on the complexity of the model. Finally, if only the surroundings of transition zones are analyzed, which can be considered the

most relevant areas for facies characterization, the improvements given by the proposed preferential solution are even more significant for the three models explored.

The rest of this chapter is organized as follows. Section 3.2 offers an overview for two important topics that are closely related with this work on: uncertainty characterization for geological models and sampling design. Section 3.3 introduces some basic information quantities that will be adopted for the formalization of the optimal sensing problem. Section 3.4 presents the general problem of optimal sampling (OS) and Section 3.5 extends this formulation to its adaptive sensing variation that is essential to apply it in the context of conditional simulations. Section 3.6 presents the preliminary study of one-dimensional Markov chain and Section 3.7 addresses the main application in MPS. Finally, conclusions and final remarks are presented in Section 3.8. Some complementary material is relegated to the Appendices.

3.2 Related Works on Geological Uncertainty and Sampling Design

The analysis and quantification of uncertainty in geological models have been systematically addressed by the geo-scientific community (Goovaerts, 2001; Gutjahr, 1991; Scheidt and Caers, 2009; Wellmann and Regenauer-Lieb, 2012). In this context, the Shannon entropy has been proposed to visualize and analyze spatial uncertainties in the context of structural maps and complex three-dimensional spatial geological models (Goodchild et al, 1994; Wellmann and Regenauer-Lieb, 2012; Wellmann et al, 2010). Preliminary association between geostatistics and information theory can be found in (Peschel and Mokošch, 1991), where the relationship between information entropy and the estimation variance of geostatistical methods is established under a joint Gaussian assumption for the model. More recently, (Wellmann, 2013) proposed the adoption of conditional entropy and mutual information to quantify spatial dependencies and spatial uncertainty reduction (attributed to the information) of a measurement (a drilling well) considering a typical geological model. In addition, (Schweizer et al, 2017) considers the entropy as a measure to assess structural uncertainties by the comparison of multiple model interpretations and the subsequent track of changes across built models. Here the entropy is used to understand how a three-dimensional geological model evolves as new data of increasing complexity is integrated in the analysis.

Considering the problem of optimal sensor placement (or sampling design), this is a fundamental problem that has been addressed in numerous contexts. For instance a classical task in signal processing is the sampling problem (Eldar, 2015), where the objective is to recover the signal from a under-sampling set of measurements (typically linear measurements) (Candes et al, 2006c; Donoho, 2006; Founcart and Lai, 2009; Vershynin, 2012). In this context, the optimal sampling design has been systematically addressed for the objective of full signal recovery (or perfect reconstruction) (Boyko et al, 2014; Candes, 2008; Candes et al, 2006b) and the problem of estimating the signal using a mean square error criterion (Bui et al, 2015; Cohen et al, 2009). On this area, compressed sensing (CS)

provides concrete results concerning the optimal sampling scheme, the reconstruction algorithms and the minimum number of samples needed to achieve optimal performance for signal recovery under some assumption about the structure of the signal (Baraniuk et al, 2008; Candes, 2008). Other widely explored application is in the context of sensor networks, where the problem of optimal sensor placements has been addressed systematically and the literature is rich (Bangerth et al, 2005, 2006; Guestrin et al, 2005; Krause et al, 2006, 2008a,b, 2011).

Closer to the area of application of this work, the problem of sensor placement has been addressed in the context of underground CO₂ monitoring (Magnant, 2011), geological investigation (Xu et al, 2017), waterloss detection in a water distribution network (Christodoulou et al, 2013) and kriging for continuous Gaussian fields (Abellan and Noetinger, 2010; Cressie et al, 1990; Marchant and Lark, 2007; Peschel and Mokosch, 1991; Zidek et al, 2000). In this last geostatistical context, the problem has been traditionally addressed for continuous variables by the minimization of the estimation variance coming from kriging or some of its variants (Brus and Heuvelink, 2007; Cressie et al, 1990; Gao et al, 1996; Marchant and Lark, 2007; McBratney and Webster, 1981; McBratney et al, 1981; Olea, 1984; Vasát et al, 2010) or from an economic utility function of this uncertainty (Aspie and Barnes, 1990). One line of work addressed the problem by using heuristics or by splitting it into a set of sequential decisions. Simulated annealing (Christakos and Killam, 1993; van Groenigen et al, 1999; Norrena and Deutsch, 2002), genetic algorithms (Bittencourt and Horne, 1997) or particle swarm optimization (Afshari et al, 2014), numerical sparse approaches (Magnant, 2011) are among the typical approaches to overcome the complexity of the problem. On the other hand, (Abellan and Noetinger, 2010) proposed the use of information theory to formulate the optimal sampling design where the utility function selects the points that maximize the difference between the prior and posterior density of the model parameters given the measurements. They evaluate this approach in the context of kriging of a full Gaussian model where the Kullback-Leibler divergence (KLD) is used to measure the discrepancy between the prior and posterior density. In the context of geological prediction, the principle of maximum entropy sampling under a Markov assumption for the model was proposed for the disposition of observations points in (Xu et al, 2017) and the same principle was also presented for the context of longitudinal sensing (meaning acoustic, pressure and flow sensors) for water-loss detection in a water distribution network in (Christodoulou et al, 2013).

The extension of some of the information theoretic formulations for sampling design presented above (like the one proposed for kriging in the context of a multivariate Gaussian assumption in (Abellan and Noetinger, 2010), or the maximum entropy principle proposed under the Markov assumptions in (Xu et al, 2017) for geological investigation) to the problem of conditional simulations of categorical random fields using multiple-point statistics to the best of the knowledge of the authors has not been addressed in the literature and constitute the objective of the rest of the exposition.

3.3 Preliminaries: Notation and Basic Concepts

A random variable (and a random vector) will be denoted by capital letters, for example $X(w) : (\Omega, \mathbb{P}) \rightarrow \mathcal{A}$ where μ_X denotes the distribution of X in the alphabet \mathcal{A} . Focusing on the finite alphabet case, $\mu_X(x)$ for all $x \in \mathcal{A}$ will be a short-hand for the probability mass function (pmf). When considering a joint vector (X, Y) in $\mathcal{A} \times \mathcal{B}$, with $\mu_{X,Y}$ denoting its joint distribution and $\mu_{X|Y}(\cdot|y)$ the conditional distribution of X given $Y = y$ in \mathcal{A} . Finally $\mathcal{P}(\mathcal{A})$ denotes the collection of probabilities defined in \mathcal{A} .

3.3.1 Information Measures

Let X be a random variable with values in a finite alphabet \mathcal{A} and probability measure $\mu_X \in \mathcal{P}(\mathcal{A})$. The Shannon entropy of X (Shannon, 1948) (or alternatively of μ_X) is a measure of the uncertainty or non-resolvability of X given by

$$H(X) \equiv \mathcal{H}(\mu_X) \equiv - \sum_{x \in \mathcal{A}} \mu_X(x) \log_2 \mu_X(x). \quad (3.1)$$

This object has been widely studied on information theory, where the convention from general theoretical formulation is to use the integer 2 as the base of the logarithm. The Shannon entropy offers fundamental performance bounds to many coding problems (Cover and Thomas, 2006). Other important concepts are conditional entropy and mutual information (Cover and Thomas, 2006). Consider two random variables X and Y with values on finite alphabets \mathcal{A} and \mathcal{B} , respectively, and joint distribution $\mu_{X,Y}$ in $\mathcal{P}(\mathcal{A} \times \mathcal{B})$. Then, the entropy of X given $Y = y \in \mathcal{B}$ from (3.1) is given by

$$H(X|Y = y) \equiv \mathcal{H}(\mu_{X|Y}(\cdot|y)) \equiv - \sum_{x \in \mathcal{A}} \mu_{X|Y}(x|y) \log_2 \mu_{X|Y}(x|y). \quad (3.2)$$

The average value of $H(X|Y = y)$ (Cover and Thomas, 2006) with respect to the statistics of Y is defined as the conditional entropy of X given Y

$$H(X|Y) \equiv \sum_{y \in \mathcal{B}} \mu_Y(y) H(X|Y = y). \quad (3.3)$$

From Jensen's inequality (Cover and Thomas, 2006), it is well-known that $H(X|Y) \leq H(X)$ where the equality is achieved, if and only if, X and Y are independent. Therefore, the basic principle is that conditioning always reduces the uncertainty. In information theory, the reduction of uncertainty is attributed to information, and consequently, the mutual information of X and Y is defined precisely by

$$I(X; Y) \equiv H(X) - H(X|Y) \geq 0. \quad (3.4)$$

Then $I(X; Y)$ is the average reduction of uncertainty of X by the process of sensing (measuring) values of Y . The mutual information is a symmetric measure (Cover and Thomas, 2006), that is $H(Y) - H(Y|X) = H(X) - H(X|Y)$. Thus, the alternative interpretation can be used by changing the role between X and Y .

3.3.2 Entropy as an Indicator of Simulation Complexity

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 (see App. A.1). This in the sense that the minimum number of n -block samples needed to represent almost all the probability of the phenomenon scales like $\approx 2^{n \cdot H(X)}$, indicating an exponential rate that is fully determined by $H(X)$. From this angle, the task of simulating X_1 is more complex than simulating X_2 if $H(X_1) > H(X_2)$. The derivation of this result comes from a natural connection between the task of i.i.d. simulation and the problem of almost lossless source coding (Cover and Thomas, 2006) and, in particular, a connection with the celebrated concept of typical sequences introduced by Shannon (Shannon, 1948) to prove numerous coding theorems. For completeness, this material is presented in App. A.1 of this thesis.

3.3.3 A Basic Sensing Decision Problem

Here an illustration of a binary sensing problem is posted for the task of i.i.d. simulation that is helpful to understand the formulation of the sensing problem presented in Sects. 3.4 and 3.5. Let X denote a finite alphabet random variable (with values in \mathcal{A}) to be characterized through i.i.d. samples. Before performing the simulation, one out of two possible finite alphabet random variables Y_1 and Y_2 with values in the same alphabet \mathcal{B} can be measured. Both random variables, Y_1 and Y_2 , have a joint distribution with the target random variable X , and the problem is to choose optimally one of them before doing the i.i.d. simulations of X . Here the joint distribution μ_{X,Y_j} was assumed as non-trivial in the sense that $\mu_{X,Y_j} \neq \mu_X \times \mu_{Y_j}$, in other words, that $I(X; Y_j) > 0$ from (3.4).

From the previous section, the complexity of the i.i.d. simulation task is proportional to the entropy of its marginal distribution. Therefore if $Y_j = y \in \mathcal{B}$ is observed, the complexity of creating i.i.d. samples of X , is given by

$$H(X|Y_j = y) = \mathcal{H}(\mu_{X|Y_j}(\cdot|y)) = - \sum_{x \in \mathcal{A}} \mu_{X|Y_j}(x|y) \log_2 \mu_{X|Y_j}(x|y). \quad (3.5)$$

As the specific value that Y_j will take is a priori unknown, the average complexity is considered with respect to the probability of Y_j , that is,

$$H(X|Y_j) = \sum_{y \in \mathcal{B}} \mu_{Y_j}(y) \cdot H(X|Y_j = y). \quad (3.6)$$

Then the minimum cost decision for simulating X conditioned on a sensed variable taken from the set $\{Y_1, Y_2\}$ is obtained from

$$j^* = \arg \min_{j \in \{1,2\}} H(X|Y_j) \quad (3.7)$$

$$= \arg \max_{j \in \{1,2\}} H(X) - H(X|Y_j) \quad (3.8)$$

$$= \arg \max_{j \in \{1,2\}} I(X; Y_j). \quad (3.9)$$

The solution in (3.7) can be interpreted as selecting the sensed variable that minimizes the posterior uncertainty (in average) of X after taking the measurement in (3.7). Alternatively, (3.9) is equivalent to the problem of selecting the sensed variable that has the highest mutual information with X or the highest reduction between the prior and the posterior uncertainty in (3.8). In fact, if comparing the complexity of the task of simulating X with and without the ability of optimally sensing over the set $\{Y_1, Y_2\}$, there is a reduction in complexity quantified by $H(X) - H(X|Y_{j^*})$ which is precisely $I(X; Y_{j^*})$, the highest point-wise mutual information between X and the set $\{Y_1, Y_2\}$ stated in (3.9).

3.4 Sampling Strategies for Uncertainty Reduction

In this section, the decision principle in Sect. 3.3.3 is extended to the case of characterizing a random field from a finite collection of measurements or pixel-based samples. In this work, only the squared two dimensional problem is addressed for a sake of simplicity, but no limitations exist to extend it to more complex problems (e.g., three-dimensional rectangular grids).

3.4.1 Problem Setting

The object to be characterized is a random field corresponding to a collection of finite alphabet random variables. $\bar{X} = \{X_{i,j} : (i, j) \in [M] \times [M]\}$ where $[M] \equiv \{1, \dots, M\}$. For every position in the array (i, j) , $X_{i,j}$ is a random variable with values in a finite alphabet \mathcal{A} and marginal probability $\mu_{X_{i,j}} \in \mathcal{P}(\mathcal{A})$. The collection \bar{X} is equipped with its joint probability denoted by $\mu_{\bar{X}}$ in $\mathcal{P}(\mathcal{A}^{M^2})$.

Following the task in Sect. 3.3.3, 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 $\mathbf{F}_k \equiv \{f : \{1, \dots, k\} \rightarrow [M] \times [M]\}$ 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 $\mathcal{I}_f \equiv \{f(1), f(2), \dots, f(k)\} \subset [M] \times [M]$. In particular for $f \in \mathbf{F}_k$, let

$$X_f \equiv (X_{f(1)}, X_{f(2)}, \dots, X_{f(k)}), \quad (3.10)$$

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

$$\hat{X}_f \equiv (X_{i,j} : (i, j) \in [M] \times [M] \setminus \{f(1), f(2), \dots, f(k)\}) \quad (3.11)$$

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

position \hat{X}_f is given by (see Sect. 3.3.2)

$$\begin{aligned} H(\hat{X}_f|X_f = \bar{x}) &= \mathcal{H}(\mu_{\hat{X}_f|X_f}(\cdot|\bar{x})) \\ &= - \sum_{\bar{y}=(y_1, \dots, y_{M^2-k}) \in \mathcal{A}^{M^2-k}} \mu_{\hat{X}_f|X_f}(\bar{y}|\bar{x}) \cdot \log_2 \mu_{\hat{X}_f|X_f}(\bar{y}|\bar{x}). \end{aligned} \quad (3.12)$$

In general, there is no access to the specific realization of X_f when evaluating the rule. Consequently, the cost associated to f is the average posterior uncertainty with respect to the statistic of X_f given by

$$H(\hat{X}_f|X_f) = \sum_{\bar{x}=(x_1, \dots, x_k) \in \mathcal{A}^k} \mu_{X_f}(\bar{x}) \cdot H(\hat{X}_f|X_f = \bar{x}). \quad (3.13)$$

Then, the minimum cost decision rule associated to characterize the field with k sensed positions is given by

$$f_k^* = \arg \min_{f \in \mathbf{F}_k} H(\hat{X}_f|X_f). \quad (3.14)$$

3.4.2 The Equivalent Maximum Information Strategy

As mentioned in previous sections, uncertainty of the field after measuring translates into information. Therefore, there is a meaningful interpretation of (3.14) as a problem of maximum information extraction. The amount of information that provides a decision rule $f \in \mathbf{F}_k$ about \bar{X} , is defined as the reduction of the uncertainty induced in \bar{X} after taking the measurements in average. Thus, the information of f to resolve \bar{X} is

$$\begin{aligned} I(f) &\equiv I(\bar{X}; X_f) = H(\bar{X}) - H(\bar{X}|X_f) \\ &= H(\bar{X}) - H(\hat{X}_f|X_f), \end{aligned} \quad (3.15)$$

where $H(\bar{X})$ is the Shannon entropy of the entire field (before the measurements) or the a priori uncertainty of \bar{X} given by

$$H(\bar{X}) = \mathcal{H}(\mu_{\bar{X}}) = - \sum_{\bar{x}=(x_{i,j})_{(i,j) \in [M] \times [M]} \in \mathcal{A}^{M^2}} \mu_{\bar{X}}(\bar{x}) \cdot \log_2 \mu_{\bar{X}}(\bar{x}), \quad (3.16)$$

and the last equality in (3.15) comes from the chain rule of the entropy (Cover and Thomas, 2006) and the fact that $\bar{X} = (X_f, \hat{X}_f)$. By definition, $I(f)$ is a particular case of the mutual information (Cover and Thomas, 2006), which implies that $I(f) \geq 0$ and $I(f) \leq H(\bar{X})$. An interesting case to consider is when $I(f) = H(\bar{X})$. This implies that $H(\hat{X}_f|X_f) = 0$, which is equivalent to say that \hat{X}_f is a deterministic function of X_f (Cover and Thomas, 2006) and, consequently, the sensing rule f perfectly resolves \bar{X} with no remaining uncertainty.

Then, the optimal rule stated in Eq.(3.14) can be posted from (3.15) as

$$f_k^* = \arg \max_{f \in \mathbf{F}_k} I(f), \quad (3.17)$$

which has the nice interpretation of maximizing the information to resolve \bar{X} with k measurements.

Another interpretation of the decision problem in (3.14) can be provided using the chain rule of the entropy (Cover and Thomas, 2006). From the chain rule, for any $f \in \mathbf{F}_k$ the joint entropy $H(\bar{X})$ can be decomposed as the sum of the marginal entropy of the sensed pixels $H(X_f)$ and the conditional entropy $H(\hat{X}_f|X_f)$. Consequently from (3.15),

$$f_k^* = \arg \max_{f \in \mathbf{F}_k} H(X_f), \quad (3.18)$$

which is the problem of choosing the k positions that jointly lead to the highest a priori (before the measurements) entropy.

Remark 1 Summarizing, the problem of minimizing the posterior Shannon entropy after taking the measurements in Eq. (3.14) is equivalent to the problem of finding the k measurements that maximize the information to resolve \bar{X} in Eq. (3.17), which is also equivalent to finding the subset of k measurements that maximize the a priori uncertainty before taking the measurements in Eq. (3.18).

Finally, after solving the optimal sensing rule of size k , the resulting posterior uncertainty is given by

$$H(\hat{X}_{f_k^*}|X_{f_k^*}). \quad (3.19)$$

On the other hand, the reduction on the uncertainty (prior minus posterior entropy), attributed to the information of f_k^* to resolve \bar{X} , is given by

$$I(f_k^*) = H(\bar{X}) - H(\hat{X}_{f_k^*}|X_{f_k^*}) = H(X_{f_k^*}) \geq 0. \quad (3.20)$$

3.4.3 The Iterative Sequential Rule

The optimal sensing rule in its three presentations in (3.14), (3.17) and (3.18) is a combinatorial problem and, consequently, impractical for relatively large fields. In this section, the sequential (non-adaptive) maximum information scheme (SMIS) is proposed as an iterative solution based on the principle of one-step-ahead sensing. The idea is to construct a sub-optimal sensing rule in an incremental fashion to reduce the complexity of the decision task to something polynomial in the size of the problem and, therefore, that can be implemented in practice.

In this context, for $k = 1$, the optimal rule reduces to finding one position in the array solution of

$$(i_1^*, j_1^*) = \arg \min_{(i,j) \in [M] \times [M]} H((X_{p,s} : (p,s) \neq (i,j)) | X_{i,j}) \quad (3.21)$$

$$= \arg \max_{(i,j) \in [M] \times [M]} H(\bar{X}) - H((X_{p,s} : (p,s) \neq (i,j)) | X_{i,j}) \quad (3.22)$$

$$= \arg \max_{(i,j) \in [M] \times [M]} H(X_{i,j}), \quad (3.23)$$

where (3.21) comes from (3.14), (3.22) from (3.17) and (3.23) from (3.18).

For $k = 2$, the principle of the one-step ahead approach fixes (i_1^*, j_1^*) and finds the next position in the array as the solution of

$$(i_2^*, j_2^*) = \arg \min_{(i,j) \in [M] \times [M] \setminus \{(i_1^*, j_1^*)\}} [H((X_{p,s} : (p,s) \neq (i,j), (p,s) \neq (i_1^*, j_1^*)) | X_{i,j}, X_{i_1^*, j_1^*})] \quad (3.24)$$

$$= \arg \max_{(i,j) \in [M] \times [M] \setminus \{(i_1^*, j_1^*)\}} [H((X_{p,s} : (p,s) \neq (i_1^*, j_1^*)) | X_{i_1^*, j_1^*}) - H((X_{p,s} : (p,s) \neq (i,j), (p,s) \neq (i_1^*, j_1^*)) | X_{i,j}, X_{i_1^*, j_1^*})] \quad (3.25)$$

$$= \arg \max_{(i,j) \in [M] \times [M] \setminus \{(i_1^*, j_1^*)\}} H(X_{i,j} | X_{i_1^*, j_1^*}). \quad (3.26)$$

Again, the first problem minimizes the posterior uncertainty after taking the next measurement, the second maximizes the information gain of the next measurement, and the last finds the point that maximizes the a priori uncertainty (conditioning to previous data, in this case $X_{i_1^*, j_1^*}$), which is the simplest principle to implement in practice.

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,

$$(i_k^*, j_k^*) = \arg \max_{(i,j) \in [M] \times [M] \setminus \{(i_l^*, j_l^*) : l=1, \dots, k-1\}} H(X_{i,j} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-1}^*, j_{k-1}^*}). \quad (3.27)$$

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

$$\tilde{f}_k^*(1) = (i_1^*, j_1^*), \tilde{f}_k^*(2) = (i_2^*, j_2^*), \dots, \text{ and } \tilde{f}_k^*(k) = (i_k^*, j_k^*). \quad (3.28)$$

Figure 3.1 illustrates the inputs needed to solve (3.27), that is, the statistical model and previous positions, and Fig. 3.2 the sequential set of problems and their interplay to obtain $\tilde{f}_k^*(\cdot)$ in (3.28).

In general, the rule in Eq. (3.27) can reach its maximum at multiple locations, giving rise to multiple feasible solutions. In this context, a random selection from feasible locations will be used in practice. This is made in order to exclusively use entropy metrics as the source of preferential sampling.

Concerning the information of this sequential rule, the following can be stated:

PROPOSITION 1 The information of \tilde{f}_k^* to resolve the field $\{X_{i,j}\}$ (using the definition in (3.15)) is given by

$$\begin{aligned} I(\tilde{f}_k^*) &= H(\bar{X}) - H(\bar{X} | X_{\tilde{f}_k^*}) = H(\bar{X}) - H(\hat{X}_{\tilde{f}_k^*} | X_{\tilde{f}_k^*}) \\ &= H(X_{i_1^*, j_1^*}, \dots, X_{i_{k-1}^*, j_{k-1}^*}, X_{i_k^*, j_k^*}) \\ &= H(X_{i_1^*, j_1^*}) + H(X_{i_2^*, j_2^*} | X_{i_1^*, j_1^*}) + \dots + H(X_{i_k^*, j_k^*} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-1}^*, j_{k-1}^*}), \end{aligned} \quad (3.29)$$

and, consequently, the information gain of this approach is additive in the sense that

$$I(\tilde{f}_k^*) - I(\tilde{f}_{k-1}^*) = H(X_{i_k^*, j_k^*} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-1}^*, j_{k-1}^*}) \geq 0. \quad (3.30)$$

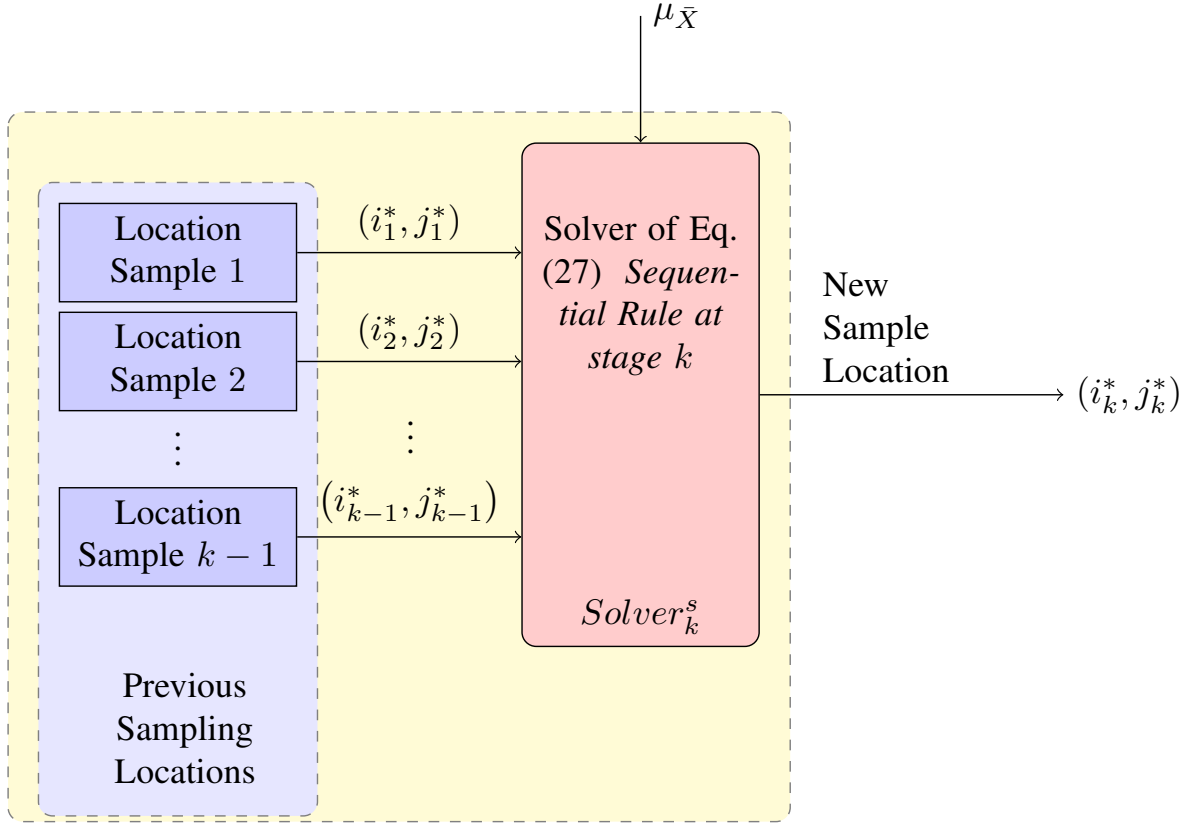


Figure 3.1: Diagram of the inputs and statistical information (model) needed to solve the sequential rule in Eq. (3.27)

(The proof is presented in App. A.2)

Remark 2 Note that the information gain from $k-1$ to k measurements in (3.30) is the solution of (3.27) in the iteration k of the sequential rule.

Finally after obtaining the optimal sequential rule, the remaining uncertainty (or posterior entropy) of the field is given by

$$H(\hat{X}_{\tilde{f}_k^*} | X_{\tilde{f}_k^*}) = H(\bar{X}) - I(\tilde{f}_k^*), \quad (3.31)$$

where $I(\tilde{f}_k^*)$ is given in Proposition 1.

Remark 3 From the definitions in (3.17) and (3.28), it is simple to show that the combinatorial rule f_k is better than the sequential rule \tilde{f}_k^* in the sense that

$$I(\tilde{f}_k^*) \leq I(f_k^*). \quad (3.32)$$

It is possible to conjecture that the information loss $(I(f_k^*) - I(\tilde{f}_k^*))_{k=1, \dots, M^2}$ is proportional to how much spatial dependency is presented in the joint distribution of field \bar{X} . To the extreme, it is simple to prove for a field with no inter-pixel dependency, in the senses that $\mu_{\bar{X}}((x_{i,j})_{(i,j) \in [M] \times [M]}) = \prod_{(i,j) \in [M] \times [M]} \mu_{X_{i,j}}(x_{i,j})$, that

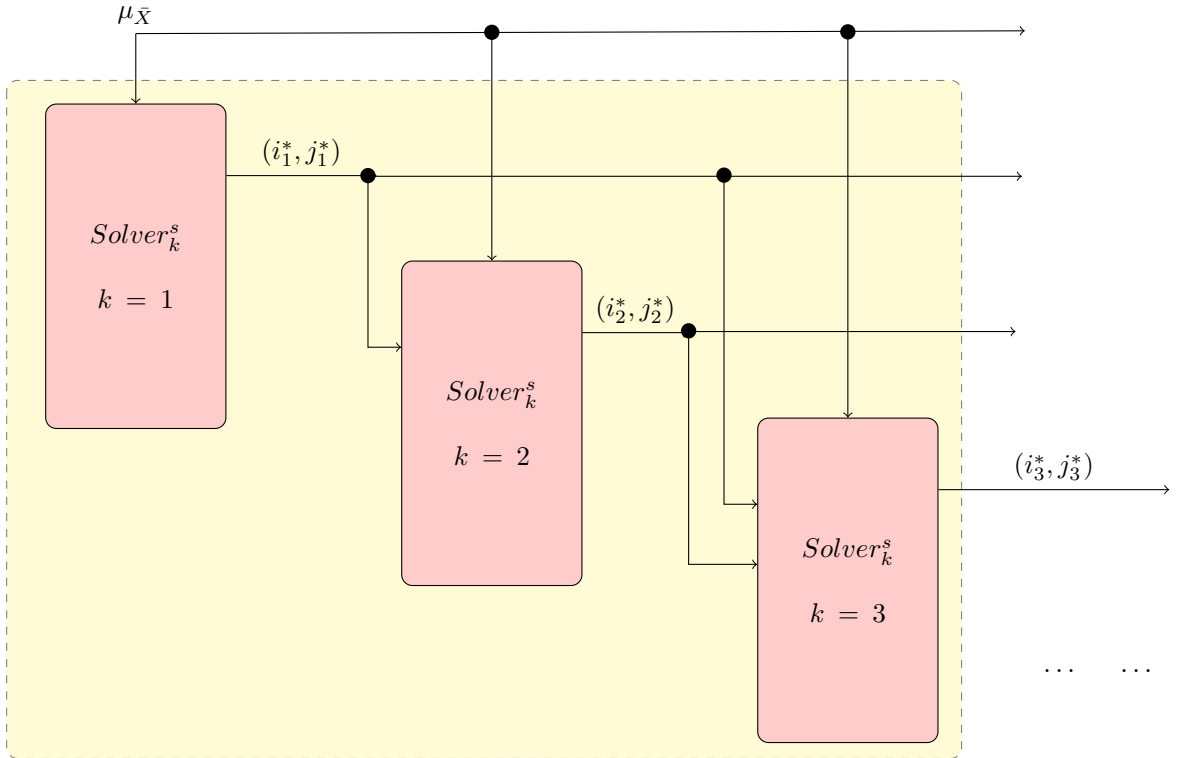


Figure 3.2: Flow diagram to determine the sampling rule $\tilde{f}_k^*(\cdot)$ in Eq. (3.28)

$$I(\tilde{f}_k^*) = I(f_k^*), \quad (3.33)$$

and the sequential solution is an optimal solution of (3.17) (see details in App. A.3).

3.5 The Adaptive Sensing Problem

In this section, the adaptive maximum information sampling (AMIS) is introduced as an adaptive sensing variation for the sequential strategy elaborated in Sect. 3.4.3. The adaptive nature of this variation comes from the practical observation that in the sequential step, for instance after deciding the $k - 1$ positions, the optimizer of the next sensing decision at stage k can have access to the actual measurements of the field taken on the previous positions. This is a realistic consideration in the context of MPS because it is the way in which the conditional simulations are performed, that is, the non-sensed pixels are simulated based on the conditional probability using the “*specific values of sensed data*”. Consequently, it is natural to consider both the previously decided positions and the measurements of the field taken on those positions when selecting the position of the next sample.

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 (3.27), an adaptive strategy

can condition on the specific values previously measured at the $k - 1$ positions. Then, assuming access to the “true data” $(x_1, \dots, x_{k-1}) \in \mathcal{A}^{k-1}$ of the image at the positions $(i_1^a, j_1^a), \dots, (i_{k-1}^a, j_{k-1}^a)$, the next position is the solution of the AMIS approach, given by

$$\begin{aligned} & (i_k^a(x_1, \dots, x_{k-1}), j_k^a(x_1, \dots, x_{k-1})) = \\ & \arg \max_{(i,j) \in [M] \times [M] \setminus \{(i_l^a, j_l^a) : l=1, \dots, k-1\}} H(X_{i,j} | X_{i_1^a, j_1^a} = x_1, \dots, X_{i_{k-1}^a, j_{k-1}^a} = x_{k-1}). \end{aligned} \quad (3.34)$$

To distinguish this solution from (3.27), the difference on the definitions on (3.2) and (3.3) should be noted. In particular, the solution in (3.34) is a function of the following set of marginal conditional distributions in $\mathcal{P}(\mathcal{A})$

$$\left\{ \mu_{X_{i,j} | X_{i_1^a, j_1^a}, \dots, X_{i_{k-1}^a, j_{k-1}^a}}(\cdot | x_1, \dots, x_{k-1}) : (i, j) \text{ non-sensed at iteration } k - 1 \right\} \quad (3.35)$$

and, consequently, of the measured data (x_1, \dots, x_{k-1}) . In other words, the decision at iteration k is a function of the sensed values of the previous positions, contrasting with the sequential approach in (3.27) that does not depend on the values of the previous samples.. Therefore this scheme is adaptive in the sense that it uses information of previously sampled data.

Remark 4 Another interpretation of this adaptive rule is thinking on the sequential decision in (3.34) as a one step prediction problem (Gray and Davisson, 2004), where the target is to predict the most informative position to resolve the field based on the probability model $\mu_{X_{\text{non-sensed-stage-}k} | X_{\text{sensed-stage-}k}}(\cdot | \text{conditioned on the sensed data at stage } k)$ in Eq.(3.35). This means that at any step of the rule, the probability model needs to be updated by incorporating the new data point, and then a one-step-ahead prediction problem is addressed with an adaptive model. Note that this is not the case of the sequential principle in Sect. 3.4.3, where the model of the field is fixed in all the stages and, consequently, this rule addresses a k -stage prediction in a sequence of one-steps to find the set of more informative positions to resolve the field in average.

3.5.1 Pseudo code to implement the Adaptive Rule

The implementation of (3.34) for $k > 1$ involves the following basic steps:

1. Having the joint distribution of the field $\mu_{\bar{X}}$, the positions $(i_1^a, j_1^a), \dots, (i_{k-1}^a, j_{k-1}^a) \in [M] \times [M]$ and previous sensed values $(x_1, \dots, x_{k-1}) \in \mathcal{A}^k$,
2. Computing the conditional distributions $\mu_{X_{i,j} | X_{i_1^a, j_1^a}, \dots, X_{i_{k-1}^a, j_{k-1}^a}}(\cdot | x_1, \dots, x_{k-1}) \in \mathcal{P}(\mathcal{A})$ for each $(i, j) \in [M] \times [M] \setminus \{(i_1^a, j_1^a), \dots, (i_{k-1}^a, j_{k-1}^a)\}$,
3. Computing the entropies $h_{i,j} = \mathcal{H}(\mu_{X_{i,j} | X_{i_1^a, j_1^a}, \dots, X_{i_{k-1}^a, j_{k-1}^a}}(\cdot | x_1, \dots, x_{k-1}))$ for all $(i, j) \in [M] \times [M] \setminus \{(i_1^a, j_1^a), \dots, (i_{k-1}^a, j_{k-1}^a)\}$,
4. Solving the problem $\max_{(i,j) \in [M] \times [M] \setminus \{(i_l^a, j_l^a) : l=1, \dots, k-1\}} h_{i,j}$.

This process determines $(i_k^a(x_1, \dots, x_{k-1}), j_k^a(x_1, \dots, x_{k-1}))$. The initial condition $k = 1$ is equivalent to solve (3.21).

With this procedure, the adaptive rule of size k that has access to the information of the data $(x_1, \dots, x_{k-1}) \in \mathcal{A}^{k-1}$, denoted by $\tilde{f}_k^a(\cdot | x_1, \dots, x_{k-1}) \in \mathbf{F}_k$, can be constructed in the following way:

$$\begin{aligned} \tilde{f}_k^a(1 | x_1, \dots, x_{k-1}) &= (i_1^*, j_1^*) \rightarrow \text{(solution of (3.21))} \\ \tilde{f}_k^a(2 | x_1, \dots, x_{k-1}) &= (i_2^a(x_1), j_2^a(x_1)) \\ &\dots \\ \tilde{f}_k^a(k | x_1, \dots, x_{k-1}) &= (i_k^a(x_1, \dots, x_{k-1}), j_k^a(x_1, \dots, x_{k-1})). \end{aligned} \quad (3.36)$$

In addition, it is simple to verify that for all $k > 1$

$$\begin{aligned} \tilde{f}_k^a(1 : k | x_1, \dots, x_{k-1}) &= \\ (\tilde{f}_{k-1}^a(1 : k-1 | x_1, \dots, x_{k-2}), (i_k^a(x_1, \dots, x_{k-1}), j_k^a(x_1, \dots, x_{k-1}))), \end{aligned} \quad (3.37)$$

where $\tilde{f}_1^a(1) = (i_1^*, j_1^*)$. For clarity, Fig. 3.3 illustrates at a higher level the inputs and output relationship obtained when solving (3.34) and Fig. 3.4, the way this basic block is executed and the information flow in the process to obtain the adaptive rule $\tilde{f}_k^a(\cdot)$ in (3.37).

3.5.2 Analysis of the Information Gain

The reduction of uncertainty or the information gained to resolve the field, attributed to the adaptive decision rule that solves (3.34), given the previous positions $(i_1^a, j_1^a), \dots, (i_{k-1}^a, j_{k-1}^a)$ and their data (x_1, \dots, x_{k-1}) , at the stage k is denoted and expressed by

$$\begin{aligned} \underbrace{I(\tilde{f}_k^a, (x_1, \dots, x_{k-1}))}_{\text{information gain}} &\equiv \underbrace{H(\hat{X}_{\tilde{f}_{k-1}^a} | X_{\tilde{f}_{k-1}^a} = (x_1, \dots, x_{k-1}))}_{\text{entropy of the prior model at stage } k-1} - \\ &\quad \underbrace{H(\hat{X}_{\tilde{f}_{k-1}^a} | X_{i_k^a, j_k^a}, X_{\tilde{f}_{k-1}^a} = (x_1, \dots, x_{k-1}))}_{\text{minimum posterior entropy when selecting } (i_k^a, j_k^a) \text{ at stage } k-1}, \\ &= H(X_{i_k^a, j_k^a} | X_{i_1^a, j_1^a} = x_1, \dots, X_{i_{k-1}^a, j_{k-1}^a} = x_{k-1}) \geq 0, \end{aligned} \quad (3.38)$$

where (i_k^a, j_k^a) is a short-hand for the solution of (3.34) that is a function of (x_1, \dots, x_{k-1}) . In fact, by using the information quantities introduced in Sect. 3.3, $I(\tilde{f}_k^a, (x_1, \dots, x_{k-1}))$ 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 (3.34)) conditioned by $X_{\tilde{f}_{k-1}^a} = (x_1, \dots, x_{k-1})$.

From Remark 4 one can anticipate that the adaptive principle should perform better than the sequential principle to resolve the field, because it is using a source of information (the data) that is not considered in the pure sequential approach. To give some evidence on this, the following simple result is stated.

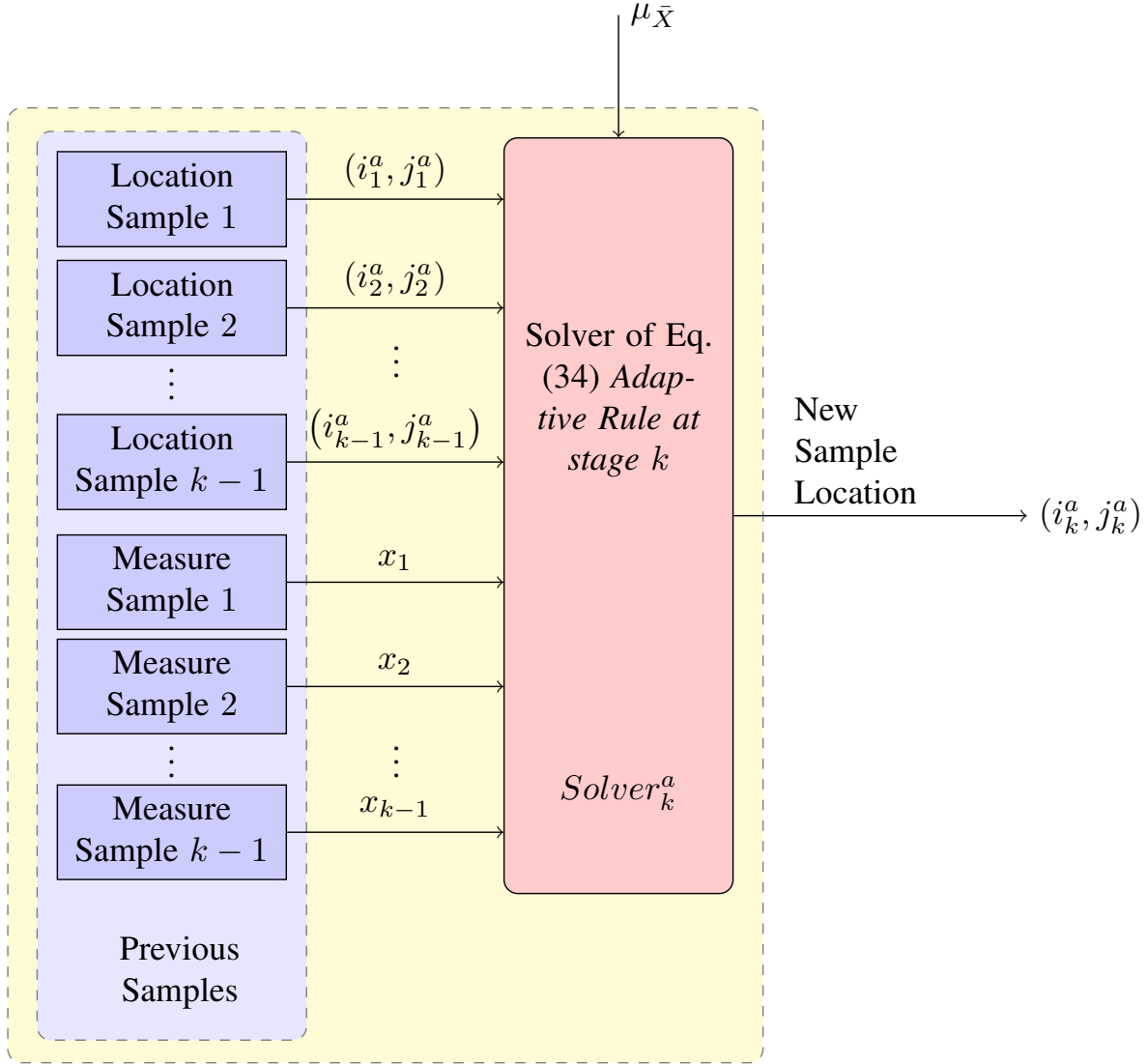


Figure 3.3: Diagram of the inputs and statistical information (model) needed to solve the adaptive rule in Eq. (3.34)

PROPOSITION 2 Consider the model $\mu_{\bar{X}} \in \mathcal{P}(\mathcal{A}^{M^2})$ of the field \bar{X} and a set of fixed positions $(i_1, j_1), \dots, (i_{k-1}, j_{k-1})$. To resolve the next position, the sequential rule that solves (i_k^*, j_k^*) from (3.27) is contrasted with the adaptive rule that solves $(i_k^a(x_1, \dots, x_{k-1}), j_k^a(x_1, \dots, x_{k-1}))$ using the side information of the values measured at positions $(i_1, j_1), \dots, (i_{k-1}, j_{k-1})$ denoted by (x_1, \dots, x_{k-1}) in (3.34). Then, for any sensed sequence $(x_1, \dots, x_{k-1}) \in \mathcal{A}^{k-1}$ having that

$$\underbrace{H(X_{i_k^*, j_k^*} | X_{i_1, j_1} = x_1, \dots, X_{i_{k-1}, j_{k-1}} = x_{k-1})}_{\text{inf. gain of the sequential rule}} \leq \underbrace{H(X_{i_k^a(x_1, \dots, x_{k-1})}, j_k^a(x_1, \dots, x_{k-1}) | X_{i_1, j_1} = x_1, \dots, X_{i_{k-1}, j_{k-1}} = x_{k-1})}_{\text{inf. gain of the adaptive rule}}. \quad (3.39)$$

Consequently, taking the average on both sides of (3.39), with respect to the realizations

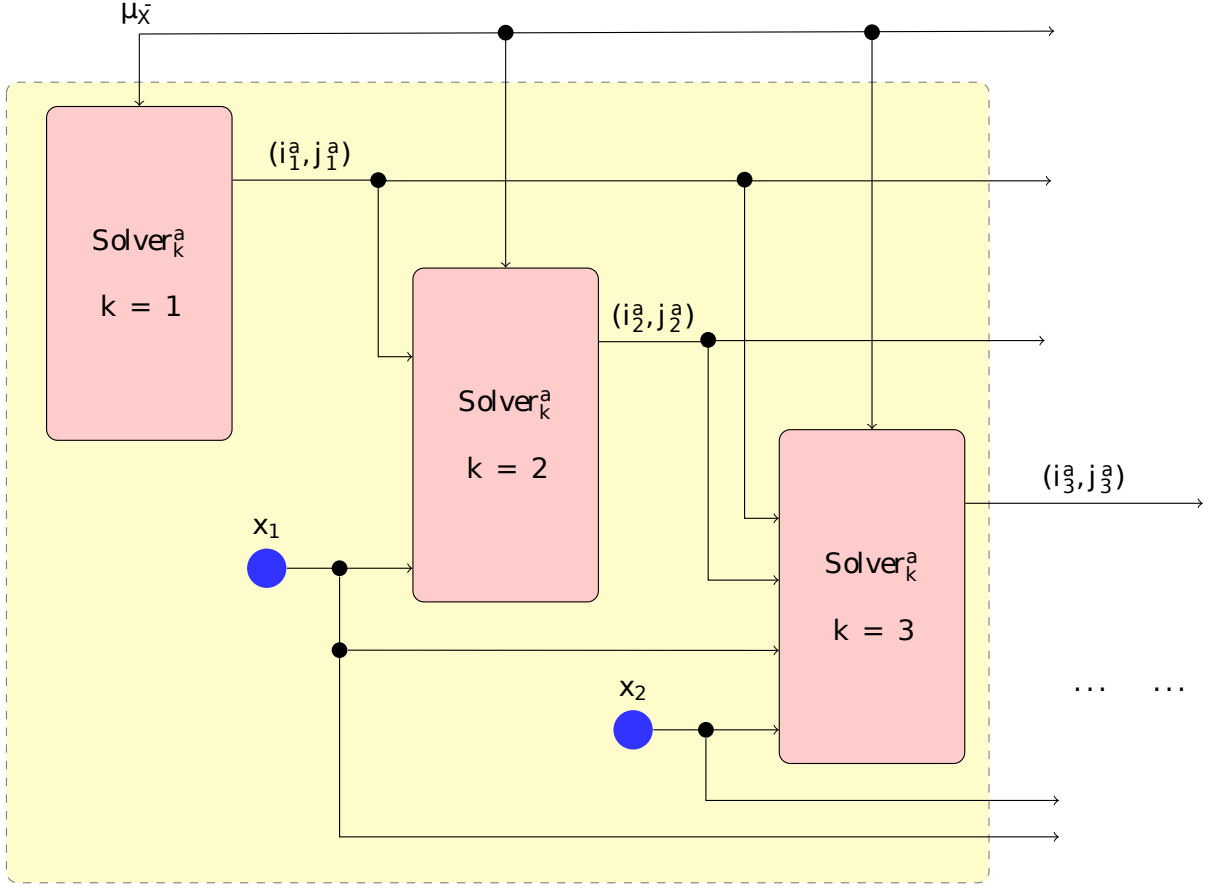


Figure 3.4: Flow diagram to determine the sampling rule in Eq. 3.37

of the sensed data $X_{(i_1, j_1)}, \dots, X_{(i_{k-1}, j_{k-1})}$ using the model $\mu_{\bar{X}}$, it provides the following identity

$$H(X_{i_k^*, j_k^*} | X_{i_1, j_1}, \dots, X_{i_{k-1}, j_{k-1}}) \leq \mathbb{E}_{Y_1, \dots, Y_{k-1}} \left\{ H(X_{i_k^a(Y_1, \dots, Y_{k-1})}, j_k^a(Y_1, \dots, Y_{k-1}) | X_{i_1, j_1} = Y_1, \dots, X_{i_{k-1}, j_{k-1}} = Y_{k-1}) \right\} \quad (3.40)$$

where $Y_1, \dots, Y_{k-1} \sim \mu_{X_{i_1, j_1}, \dots, X_{i_{k-1}, j_{k-1}}}$.

Equations (3.39) and (3.40) show the role of having access to the data in reducing the uncertainty of the field, as the information gain of the adaptive rule is better point-wise, that is, for any sequence of sensed data, and consequently in average. This result is expected and completely consistent with the fact that the adaptive rule has access to the true model to make a decision (the non-sensed variables conditioned on a specific realization of the sensed ones), while the sequential rule does not have access to the realization and, consequently, it acts over an average objective function. This fact can be observed by looking at the objective function of (3.27) and (3.34) and definition of the conditional entropy (see Sect. 3.2).

Remark 5 It is important to note that Proposition 2 compares the sequential and adaptive rules assuming that the same path has been followed until stage k . This is not the

case when inducing the sequential and adaptive rule inductively, as the paths will likely be different. The comparison of the schemes in this general scenario remains an interesting and challenging problem to be solved.

3.6 Case Study: A Binary Markov Chain

This section is devoted to the application of the proposed sequential rules in an emblematic and simple model of spatial dependency: the finite length Markov chain. Under this scenario of known statistics, the idea is to provide insights about the information-driven placement solutions as well as their performance.

A Markov chain is a simple and powerful structure to model random variables with spatial dependencies. In geology and geostatistics, Markov chains have been used to represent discrete regionalized variables such as lithologies or facies (Elfeki and Dekking, 2001; Ostroumov et al, 2005; Zhang and Li, 2008). Because of their simplicity and wide use, Markov chains offer a fantastic context to analyze the structure (in terms of spatial locations) and the performance of the optimal sampling rules elaborated in Sects. 3.4 and 3.5. Specifically, the case of $\bar{X} = (X_1, \dots, X_N)$ being a finite length binary Markov chain with values in $\mathcal{A} = \{0, 1\}$ is considered.

3.6.1 The Sequential Maximum Information Strategy (SMIS)

Beginning by applying the iterative sequential rule presented in Sect. 3.4.3, the k -stage of this problem (assuming the previous sample locations are known) is

$$i_k^* = \arg \max_{i \in [N] \setminus \{i_l^* : l=1, \dots, k-1\}} H(X_i | X_{i_1^*}, \dots, X_{i_{k-1}^*}). \quad (3.41)$$

To solve (3.41), the weak Markov property (Norris, 1997) can be used as if $i \notin f_{k-1} = \{i_1^*, \dots, i_{k-1}^*\}$ only the closest variables in $X_{\bar{f}_{k-1}} = (X_{i_j^*} : 1 \leq j \leq k-1)$ determine the conditional entropy in (3.41). More formally, if $i > \min f_{k-1}$, denoting by $l(i, k-1) \equiv \max \{t \in f_{k-1} : t < i\}$ the closest index before of i in f_{k-1} , and if $i < \max f_{k-1}$, denoting by $r(i, k-1) \equiv \max \{t \in f_{k-1} : t < i\}$ the closest index after of i in f_{k-1} , then:

PROPOSITION 3 Assuming that \bar{X} is a finite length Markov Chain and $\min f_{k-1} < i < \max f_{k-1}$, it follows that

$$H(X_i | X_{i_1^*}, \dots, X_{i_{k-1}^*}) = H(X_i | X_{l(i, k-1)}, X_{r(i, k-1)}). \quad (3.42)$$

Otherwise, if $i > \max f_{k-1}$ then $H(X_i | X_{i_1^*}, \dots, X_{i_{k-1}^*}) = H(X_i | X_{l(i, k-1)})$, and if $i < \min f_{k-1}$ then $H(X_i | X_{i_1^*}, \dots, X_{i_{k-1}^*}) = H(X_i | X_{r(i, k-1)})$.

The proof is a direct consequence of the weak Markov property and it is detailed in the App. C.5.

Proposition 3 shows the local nature of conditioning under the Markov assumption. In particular, if considering two consecutive sensed variables, $X_{i_j^*}$ and $X_{i_{j+1}^*}$, the non-sensed variables in this range ($X_h : i_j^* < h < i_{j+1}^*$) only dependent on these two extreme objects from (3.42). In addition, it is simple to verify that the conditional entropy $H(X_h | X_{i_j^*}, X_{i_{j+1}^*})$ is proportional to the distances $|h - i_j^*|$ and $|h - i_{j+1}^*|$. Thus, for a symmetric stochastic matrix, the maximal conditional entropy over ($X_h : i_j^* < h < i_{j+1}^*$) corresponds to the variable located at the center of ($X_h : i_j^* < h < i_{j+1}^*$) (see Fig. 3.5).

Extrapolating this behavior implies that the solution of (3.41) has a symmetric structure, where the optimal placement strategy locates samples uniformly over the interval $[N]$ with a specific dyadic structure, see Fig. 3.5. This sampling principle can be summarized as follows: the solution of the stage k in (3.41) is the non-sensed position that divides symmetrically the largest interval of non-sensed position induced by previous samples in f_{k-1} as it can be seen in Fig. 3.5. Only for clarity, this result is presented for a short Markov chain of 100 variables.

3.6.2 Empirical Analysis

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. The SMIS scheme $\{f_k^* : k = 1, \dots, N\}$ in (3.41) is compared with a random sampling strategy $\{f_k^{rand} : 1, \dots, N\}$ that selects N positions with a uniform distribution over $[N]$. As a performance indicator, $C(f_k)$ is introduced as the information of f_k to resolve the field in (3.15) normalized by the entropy of the entire field $H(\bar{X})$. In particular,

$$C(f_k) \equiv \frac{I(f_k)}{H(\bar{X})} = \frac{\sum_{i=2}^{k-1} H(X_{f(i)} | X_{f(i-1)}) + H(X_{f(1)})}{\sum_{i=2}^N H(X_i | X_{i-1}) + H(X_1)} \in [0, 1]. \quad (3.43)$$

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, $H(\bar{X}_{f_k} | \bar{X}_{f_k}) = 0$.

The second expression in the equation (3.43) is obtained by the fact that $I(f_k) = H(X_{f_k})$ in (3.20), the chain rule of the entropy (Cover and Thomas, 2006) and the Markov assumption, where for simplicity it is assumed that $f_k(1) < f_k(2) < \dots < f_k(k)$.

In terms of structure, Fig. 3.5 shows that SMIS has the expected dyadic structure. Remarkably, this structure is independent of the value of β and the initial distribution of X_1 , meaning that this rule is somehow universal over symmetric one-dimensional Markov processes. On the other hand, in terms of performance, Fig. 3.6 shows that when k is very small, there is no significant difference between the random scheme and SMIS, however the difference is clearer for larger k . In addition, the sampling methods show little variability between different realizations (in fact the 20 individual curves for SMIS seem to be just one apparent curve, only in the middle section of the figure is possible to appreciate a little variance in the SMIS curves). Finally, the benefit of the SMIS rule is more significant

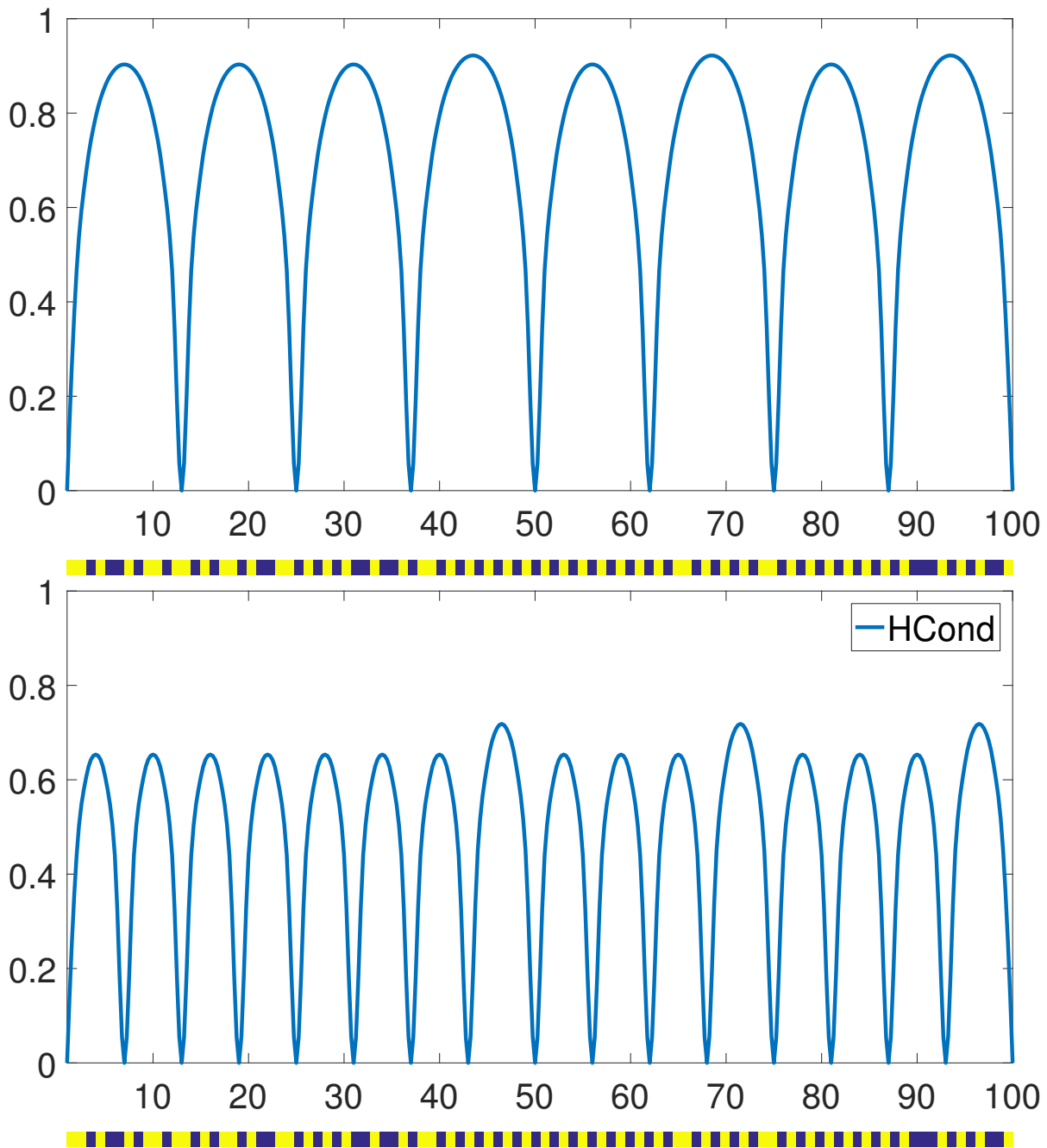


Figure 3.5: Conditional entropies given the sensed positions for a binary Markov chain. Top: after the first 10 samples. Bottom: after 18 samples. $\beta = 0.01$. Under the curves the actual realizations of the Markov chains are presented

for a scenario with more spatial dependencies, that is $\beta = 0.01$, where the field can also be resolved with a smaller number of measurements, as expected. Note that the case of $\beta = 0.5$ is closer to the i.i.d. case where in theory no difference should be expected between SMIS and a fully random approach. Note further that the difference between SIMS and random sampling resolvability capacity reaches its maximum at different sampling rates, depending on the probability of transition of the Markov processes.

In conclusion under the Markov assumption and with symmetric transition matrices, SMIS resembles a quasi-regular sampling, where the principle is to cover, as much as possible, the position space $[N]$ meaning that all positions are equally relevant. It is then remarkable than in the case of a Markov chain that models non-trivial causal dependencies, this non preferential solution is achieved with SMIS. This can be understood from the fact that the minimization is applied over the remaining uncertainty (or maximizing information of the measurements) in average, with no specific conditioning on the values of the sensed data. As presented in the next section, conditioning on the data will completely change the optimal sampling in terms of structure.

3.6.3 The Adaptive Strategy AMIS

For the adaptive sequential rule in Sect. 3.5, the Markov assumption simplifies the conditional entropy used to solve the iterative step in (3.34). In particular, if $f_{k-1}^a = \{i_1^a, \dots, i_{k-1}^a\}$ denotes the set of previous samples at stage k with known measurements $\{x_1, \dots, x_{k-1}\}$, it follows that:

PROPOSITION 4 Assuming that \bar{X} is a finite length Markov Chain and $\min f_{k-1}^a < i < \max f_{k-1}^a$, then

$$\begin{aligned} H(X_i | (X_{i_1^a}, \dots, X_{i_{k-1}^a}) = (x_1, \dots, x_{k-1})) \\ = H(X_i | X_{l(i,k-1)} = x_{l(i,k-1)}, X_{r(i,k-1)} = x_{r(i,k-1)}). \end{aligned} \quad (3.44)$$

Otherwise, if $i > \max f_{k-1}^a$ then $H(X_i | (X_{i_1^a}, \dots, X_{i_{k-1}^a}) = (x_1, \dots, x_{k-1})) =$

$H(X_i | X_{l(i,k-1)} = x_{l(i,k-1)})$, and if $i < \min f_{k-1}^a$ then $H(X_i | (X_{i_1^a}, \dots, X_{i_{k-1}^a}) = (x_1, \dots, x_{k-1})) = H(X_i | X_{r(i,k-1)} = x_{r(i,k-1)})$.

The proof is a direct consequence of the weak Markov property (Norris, 1997), and it is omitted for sake of space.

As the solution in (3.34) relies on both the Markov model and the sensed data, a deviation from the dyadic structure presented in Sect. 3.6.1 can be observed. Figure 3.7 shows the remaining conditional entropy $H(X_i | X_{i_1^a} = x_1, \dots, X_{i_{k-1}^a} = x_{k-1})$ and the sample distributions of $f_{k-1}^a = \{i_1^a, i_2^a, \dots, i_{k-1}^a\}$ where it is clear the non-regularity of the solution attributed to the spatial distribution of the data. From the performed analysis, using numerous simulations of the field, adaptation happens in the sense that the samples tend to be located around the transition zones between zero-one values. The reason is that zones around sensed positions that show transitions have more conditional uncertainty under the Markov assumption and, consequently, are more informative. This observation is very relevant because it shows that the adaptive scheme is able to put more samples on the areas close to the transitions. Importantly, this adaptive scheme provides better resolvability of the field and better recovery of non-sensed data than the sequential approach. The details are presented in App. A.4.

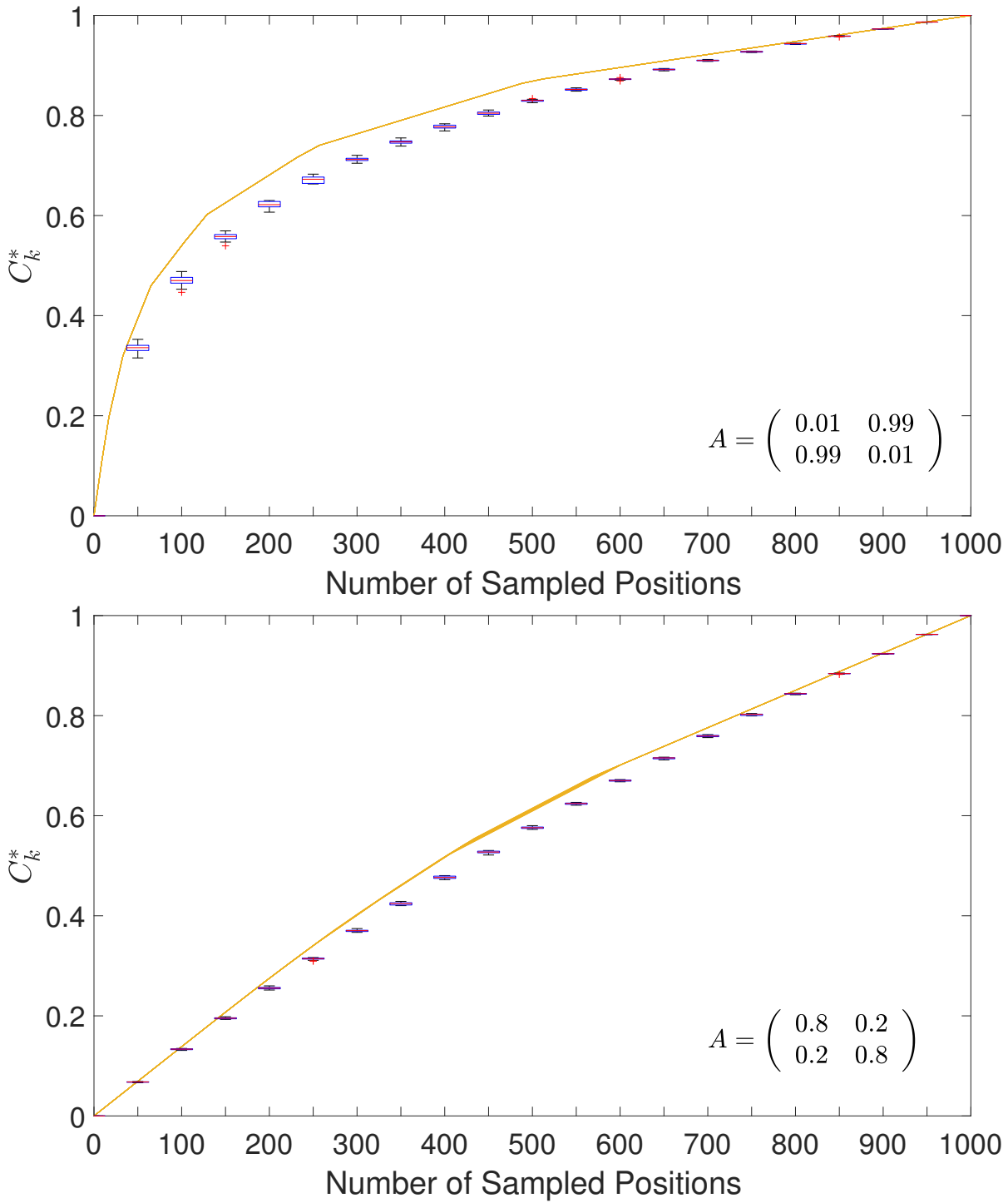


Figure 3.6: Uncertainty reduction for 20 realizations of the Markov chain. Continuous curves: SMIS; boxplots: random sampling. Top: $\beta = 0.01$, Bottom: $\beta = 0.8$

3.7 Application to Multiple-Point Simulations

In this section, the AMIS strategy in Sect. 3.5 is adopted as the main framework for preferential sampling in the context of MPS (Mariethoz and Caers, 2014). In particular,

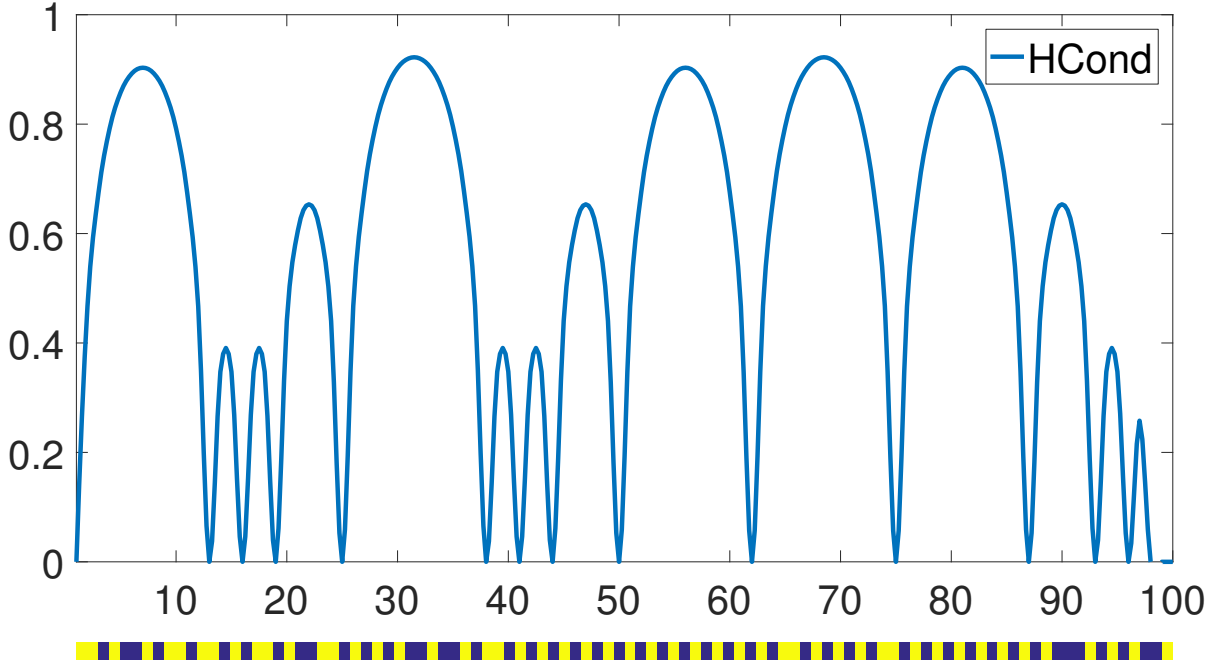


Figure 3.7: Distribution of the conditional entropy of non-sensed locations given the sensed pixels at $k = 18$, using AMIS. $\beta = 0.01$. Under the curve, the actual realization of the binary Markov chain is presented

the target is to study the applicability and effects of this sampling scheme for the problem of multiple-point conditional simulation of a binary field.

The main difficulty to execute this decision rule in MPS is the fact the joint distribution of the field $\mu_{\bar{X}} \in \mathcal{P}(\mathcal{A}^{M^2})$ is unknown (see Sect. 3.5.1). Instead, only a training image and a non-parametric approach (computing the frequency of occurrence of multiple-point patterns in the training image) are available to estimate a set of conditional probabilities of the form $\hat{\mu}_{X_{i,j}|X_S}(x_{i,j}|x_S)$ where $i, j \in [M] \times [M]$ is a position to be simulated, $S \subset [M] \times [M]$ denotes the conditioning positions (attributed to sensed data or previously simulated data), and $x_{i,j} \in \mathcal{A}$ and $x_S = (x_{i,j} : (i, j) \in S) \in \mathcal{A}^{|S|}$.

Therefore, the estimated probabilities of the form $\hat{\mu}_{X_{i,j}|X_S}(x_{i,j}|x_S)$ allow to address the basic sensing problem in (3.34). For clarity, it is important to note that MPS uses $\hat{\mu}_{X_{i,j}|X_S}(x_{i,j}|x_S)$ to simulate $X_{i,j}$ conditioned on $X_S = x_S$. Here the same model is used to select the most informative position in the sensing stage that is previous to the simulation stage. Revisiting the pseudo code presented in Sect. 3.5.1 and adopting the same notation, given a set of previous position $(i_1^a, j_1^a), \dots, (i_{k-1}^a, j_{k-1}^a)$ and their respective sensed values (x_1, \dots, x_{k-1}) , the conditional probabilities in (3.35) are estimated from the training image (the practical details of this process are relegated to Sect. 3.7.2) obtaining

$$\left\{ \hat{\mu}_{X_{i,j}|X_{i_1^a, j_1^a, \dots, i_{k-1}^a, j_{k-1}^a}}(\cdot | x_1, \dots, x_{k-1}) : (i, j) \text{ non-sensed at iteration } k-1 \right\}$$

for step 2 of this process. Then, the steps 3 and 4 follow directly to obtain $(i_k^a(x_1, \dots, x_{k-1}), j_k^a(x_1, \dots, x_{k-1}))$ in (3.34).

In addition, the sequential rule in Sect. 3.4.3 is implemented as an alternative to the adaptive rule. The problem is that this rule requires the joint distribution (see the definition in (3.3) and (3.23)), which is not directly available in MPS. A way to circumvent this issue is to use the analysis of the one-dimensional Markov model, where its non-adaptive maximum information rule can be extended to the two-dimensional case. In particular, under a generalized Markov assumption (with an isotropic conditional structure), the SMIS rule in (3.41) distributes the measurements in a structured way covering the sampling space uniformly, see Fig. 3.5. This approximation will be considered because it precisely captures one of the standard non-preferential sampling strategies used in MPS.

3.7.1 Proposed Preferential Sampling Strategy for MPS

Finally, the preferential sampling solution proposed in this work for MPS is a combination between the pure *adaptive maximum information sampling strategy* (AMIS) in (3.34) and a non-adaptive rule reminiscing of the *SMIS* strategy in (3.27) under a Markov assumption presented in Sect. 3.6.1.

Let $S_{k-1} = \{(i_l^a, j_l^a) : l = 1, \dots, k-1\}$ denote the collection of the sampled locations obtained by the proposed adaptive sampling strategy at the stage $k-1$ of the algorithm. Here, $X_{S_{k-1}} = (X_{i_1^a, j_1^a}, \dots, X_{i_{k-1}^a, j_{k-1}^a})$ corresponds to the sensed random vector indexed by S_{k-1} , and $x_{S_{k-1}} = (x_1, \dots, x_{k-1})$ 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

$$\begin{aligned} (\hat{i}_k^a(\alpha, x_{S_{k-1}}), \hat{j}_k^a(\alpha, x_{S_{k-1}})) = & \arg \max_{(i,j) \in [M] \times [M] \setminus S_{k-1}} \alpha \cdot H(X_{i,j} | X_{S_{k-1}} = x_{S_{k-1}}) \\ & + (1 - \alpha) \cdot D((i, j), S_{k-1}). \end{aligned} \quad (3.45)$$

The first term $H(X_{i,j} | X_{S_{k-1}} = x_{S_{k-1}})$ in the RHS of (3.45) is the same information term in (3.34), while the second term is a regularization term to promote uniform sampling by using a distance criterion from (i, j) to the set S_{k-1} by

$$D((i, j), S_{k-1}) = \min_{(\tilde{i}, \tilde{j}) \in S_{k-1}} d((i, j), (\tilde{i}, \tilde{j})), \quad (3.46)$$

where $d((i, j), (\tilde{i}, \tilde{j})) \equiv \sqrt{(\tilde{i} - i)^2 + (\tilde{j} - j)^2}$. In this work (3.46) is used as a regularization term, however a mean distance from $X_{i,j}$ to the set S_{k-1} , or others group distance metrics could be used as well (Scheidt and Caers, 2009).

The motivation of using this mixed solution in (3.45) instead of the pure adaptive solution ($\alpha = 1$) comes from the analysis of the Markov chain presented in App. A.4. In that context, when the number of samples is very small, structured sampling is better than adaptive sampling, but this changes as the number of samples increases, as measured by the field resolvability (conditional entropy of non-sensed position given sensed data) and the recovery of non-sensed values from sensed data. This is presented in Fig.A.1 at the App. A.4 and Fig. A.2 at the App. A.5. This suggests that there is an optimal mixture between structured and adaptive sampling which is what is proposed in (3.45).

3.7.2 Practical Considerations for the Estimation of

$$H(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}})$$

Two practical aspects about the implementation of the *RAMIS* rule in (3.45) require attention. The first comes from the effect of estimating $H(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}})$ in (3.45) for which the SNESIM algorithm (Strebelle and Zhang, 2004) is used. The second comes from the computational cost of this scheme as in every stage of (3.45) the full vector $(H(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}}))_{(i,j) \in [M] \times [M] \setminus S_{k-1}}$ has to be updated.

Regarding the first point, it is well-known that the larger the pattern the fewer the occurrences within the dimension of a training image (Remy et al, 2009). Hence large scale patterns will be poorly represented in general. To circumvent this issue SNESIM defines a searching ellipsoid characterizing the maximum pattern sizes to be considered (Huang et al, 2013; Remy et al, 2009). Here, for all models an ellipsoid with size $\frac{1}{3}$ of the image size will be used. Outside this covering zone around sensed positions, the maximum entropy value (equal to 1 for this binary problem) is adopted, corresponding to areas statistically poorly represented. Consequently at very low sampling regimes, large non-sensed areas far from sensed positions will have maximum entropy values and, consequently, the regularization term in (3.45) within these areas will dominate the *RAMIS* decision, that is, there is a structured sampling regime for the case of very small sampling rate, as expected.

On the second point, the estimation of $H(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}})$ for each $(i, j) \in [M] \times [M] \setminus S_{k-1}$ in every iteration of (3.45) ended up being extremely computationally expensive and impractical, in particular for large fields and small sampling regimes. To address this issue, the estimation of the conditional entropies is made, by computing it using SNESIM, only every K iterations at the stages $k = 1, K, 2K, \dots$. This estimation is denoted by $(\hat{H}^k(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}}))_{(i,j) \in [M] \times [M] \setminus S_{k-1}}$.

For intermediate stages, for example for $k \in \{2, \dots, K-1\}$, the values of $H(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}})$ in (3.45) are approximated, based on a neighborhood of influence obtained (a priori) from the training image and the value of this entropy estimated at stage $k-1$. In particular, for each $k = 2, \dots, K-1$ a local update of the conditional entropy is performed by the rule

$$\begin{aligned} \hat{H}^k(X_{i,j}|X_{S_{k-1}} = x_{S_{k-1}}) = & \underbrace{\hat{H}^{k-1}(X_{i,j}|X_{S_{k-2}} = x_{S_{k-2}})}_{\text{conditional entropy of the previous stage } k-1} \\ & - I(X_{i,j}; X_{\hat{i}_{k-1}^a, \hat{j}_{k-1}^a}), \end{aligned} \quad (3.47)$$

where $I(X_{i_1, j_1}; X_{i_2, j_2}) = I_{Stationary}(i_2 - i_1, j_2 - j_1)$ and $I_{Stationary}(dI, dJ)$ represents the mutual information between two positions in the field assuming stationarity and estimated from thousands of unconditional realizations generated with SNESIM. For completeness, a formal justification for the adoption of this approximation is presented in App. A.5.

To illustrate these mutual information maps, Fig. 3.8 shows the link between the spatial complexity of the training image and dependence of the obtained mutual information

maps for the three models considered in the further analysis (a single channelized model SC_1 , and two multichannel models, MC_1 and MC_2 , with different spatial complexity). From this result, training images with a higher complexity have more concentrated local mutual information dependencies than less complex training images. Thus, for models with greater spatial complexity, the effect of knowing an isolated measurement has a much more local effect in reducing global uncertainty.

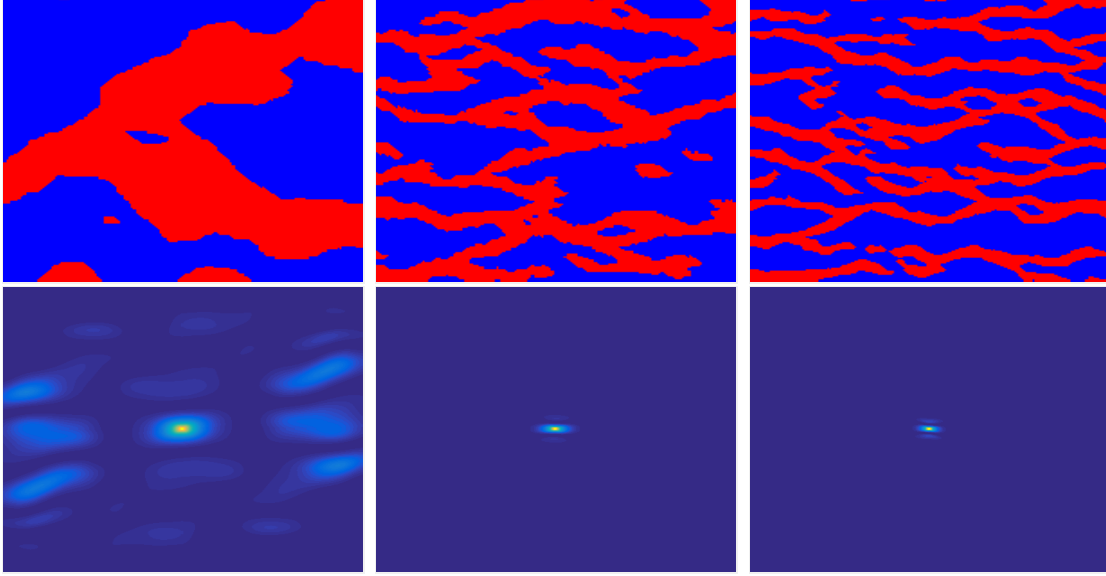


Figure 3.8: Proposed training images. Top: example of TIs, Bottom: mean MI maps estimated from 200 realizations. From left to right: Models SC_1 , MC_1 , MC_2 . Color maps, Top: Red is channel presence; Bottom: Linear from blue (low MI) to bright yellow (max. MI)

In summary, the RAMIS based decision at stage k goes as follows

$$\begin{aligned}
 (\hat{i}_k^a(\alpha, x_{S_{k-1}}), \hat{j}_k^a(\alpha, x_{S_{k-1}})) = & \arg \max_{(i,j) \in [M] \times [M] \setminus S_{k-1}} \alpha \cdot \hat{H}^k(X_{i,j} | X_{S_{k-1}}) \\
 & + (1 - \alpha) \cdot D((i, j), S_{k-1}). \tag{3.48}
 \end{aligned}$$

3.7.3 Pseudo-code of the RAMIS solution

The pseudo-code that summarizes the proposed framework is shown here in order to clarify the workflow in the MPS application.

Algorithm 1: Pseudo code that implements the adaptive framework from rules in (3.45).

```

1 Initialization
2 Choose training image  $TI$ , reference image  $\bar{X}$ , the regularization term  $\alpha$ , the batch size  $r$ ,
   and number of samples to take  $K$ 
   /* Output: set of sampled positions */
3  $f_K = \emptyset$ 
   /* Inputs and internal variables.  $\mathcal{H}^k$  Estimated Remaining entropy
   considering  $k - 1$  samples */
4  $\mathcal{H}^k = \text{ones}(\text{size}(\bar{X}))$ ,  $\mathcal{D} = \text{inf}(\text{size}(\bar{X}))$ ,  $\text{distMean} = \text{inf}$ 
5 Computation
6 for  $k \leftarrow 1$  to  $K$  do
7    $f_{k-1} \leftarrow f_K$ 
   /* Update MPS simulations from previous available samples */
8   if  $\text{Criterion\_To\_Update\_MPS\_Realizations}(k, r)$  then
9      $\mathcal{H}^k = \text{Estimated\_Entropy\_from\_MPS\_}\&\_ \text{Samples}(TI, X_{f_{k-1}}, \text{size}(\bar{X}), f_{k-1})$ 
10     $\mathcal{D} = \text{Estimated\_Distances\_From\_Sampled\_Locations}(\text{size}(\bar{X}), f_{k-1})$ 
11     $\text{distMean} = \text{Mean}(\mathcal{D})$ 
12   else
13      $\mathcal{H}^k = \mathcal{H}^k - \text{Local\_Mutual\_Information}_{\text{empirical}}(f_{k-1}(k - 1))$ 
14      $\mathcal{D} = \mathcal{D} \cdot \text{Radial\_Attenuation\_Centred\_At\_Location}(\text{size}(\bar{X}), \text{distMean}, k - 1)$ 
   /* Set of current available position to sample. The complement set of
   the previously sampled locations */
15    $\hat{f}_{k-1} = \text{Complement\_Set\_Of\_F}(f_{k-1})$ 
   /* Regularized Criterion by Mixing of Entropy and Distance */
16    $\mathcal{M} = \alpha \cdot \mathcal{H}^k + (1.0 - \alpha) \cdot \mathcal{D}$ 
   /* Choose a location randomly from the set of non sampled locations with
   maximal value for the objective function */
17    $f_K(k) = \text{Select\_Random\_Location\_From\_Maximum\_Criterion}(\mathcal{M}_{\hat{f}_{k-1}})$ 

```

3.7.4 Experimental Setting and Validation

The remaining sections evaluate the performance of the RAMIS sampling framework in (3.48) in terms of resolvability of the field, by the conditional entropy of non-sensed position given the sensed values, and also in terms of the average error between the true field and the simulations obtained with MPS across different sampling rates and facies models.

For that, 3 channelized 2-D binary facies models are studied with different levels of spatial complexity: one single channel and two multi-channel structures with predominantly horizontal orientation. These channelized structures were obtained by unconditional simulations using the geostatistical software SGeMS by the SNESIM algorithm (Huang et al, 2013; Remy et al, 2009) providing a synthetic scenario of study. The database was built

from channelized 2-D images of size 200×200 . For the analysis, from the synthetic realizations, some images are considered as ground truth for testing while thousands of others are used as training images for statistical estimation of the fields.

For the proposed adaptive approach that solves (3.48), the entropies of non-sensed values were estimated by using the empirical distribution from a set of 600 MPS conditional simulations.

For performance evaluation two metrics were considered: resolvability and simulation error. Thus, considering a rule f of size k that has sensed the positions $S_k \subset [M] \times [M]$ with values $x_{S_k} \in \{0, 1\}^k$, the resolvability of f is the average conditional entropy over the non-sensed positions, given by

$$\begin{aligned} \mathcal{R}(f, x_{S_k}) &\equiv \text{average}(H(X_{i,j}|X_{S_k} = x_{S_k})_{(i,j) \in [M] \times [M] \setminus S_k}) \\ &= \frac{1}{M^2 - |S_k|} \sum_{(i,j) \in [M] \times [M] \setminus S_k} H(X_{i,j}|X_{S_k} = x_{S_k}) \end{aligned} \quad (3.49)$$

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

$$\begin{aligned} \mathcal{E}(f, (x_{i,j})_{i,j}) &\equiv \\ &\frac{1}{M^2 - |S_{k-1}|} \sum_{(i,j) \in [M] \times [M] \setminus S_k} \mathbb{E}_{X_{i,j}} \{ (x_{i,j} - X_{i,j})^2 | X_{S_k} = x_{S_k} \} \\ &= \frac{1}{M^2 - |S_{k-1}|} \sum_{(i,j) \in [M] \times [M] \setminus S_k} \mathbb{E}_{X_{i,j}} \{ \mathbf{1}_{X_{i,j} \neq x_{i,j}} | X_{S_k} = x_{S_k} \} \\ &= \frac{1}{M^2 - |S_{k-1}|} \sum_{(i,j) \in [M] \times [M] \setminus S_k} \underbrace{\mathbb{P} \{ X_{i,j} \neq x_{i,j} | X_{S_k} = x_{S_k} \}}_{\text{Conditional Error Probability}}. \end{aligned} \quad (3.50)$$

From (3.50), $\mathcal{E}(f, (x_{i,j})_{i,j})$ corresponds to the average frequency of error in detecting the true non-sensed value (oracle) with the values simulated with MPS, over the non-sensed locations.

3.7.5 Selecting the Regularization Parameter α

The regularization parameter $\alpha \in (0, 1)$ in (3.48) offers a way of controlling the tradeoff between adaptive and non-adaptive sequential sampling. Fig. 3.9 depicts two maps that show the information and regularization terms (see the right hand side of (3.48)). Remarkably as expected, the information term shows preferential locations around the transitions and, consequently, through (3.48) adaptation to the underlying field when $\alpha > 0$. Furthermore, the regularization term offers a non-adaptive type of regular pattern covering uniformly the space $[M] \times [M]$. Therefore, the solution $\alpha = 0$ reduces to the structured sampling (see Fig. 3.9, right).

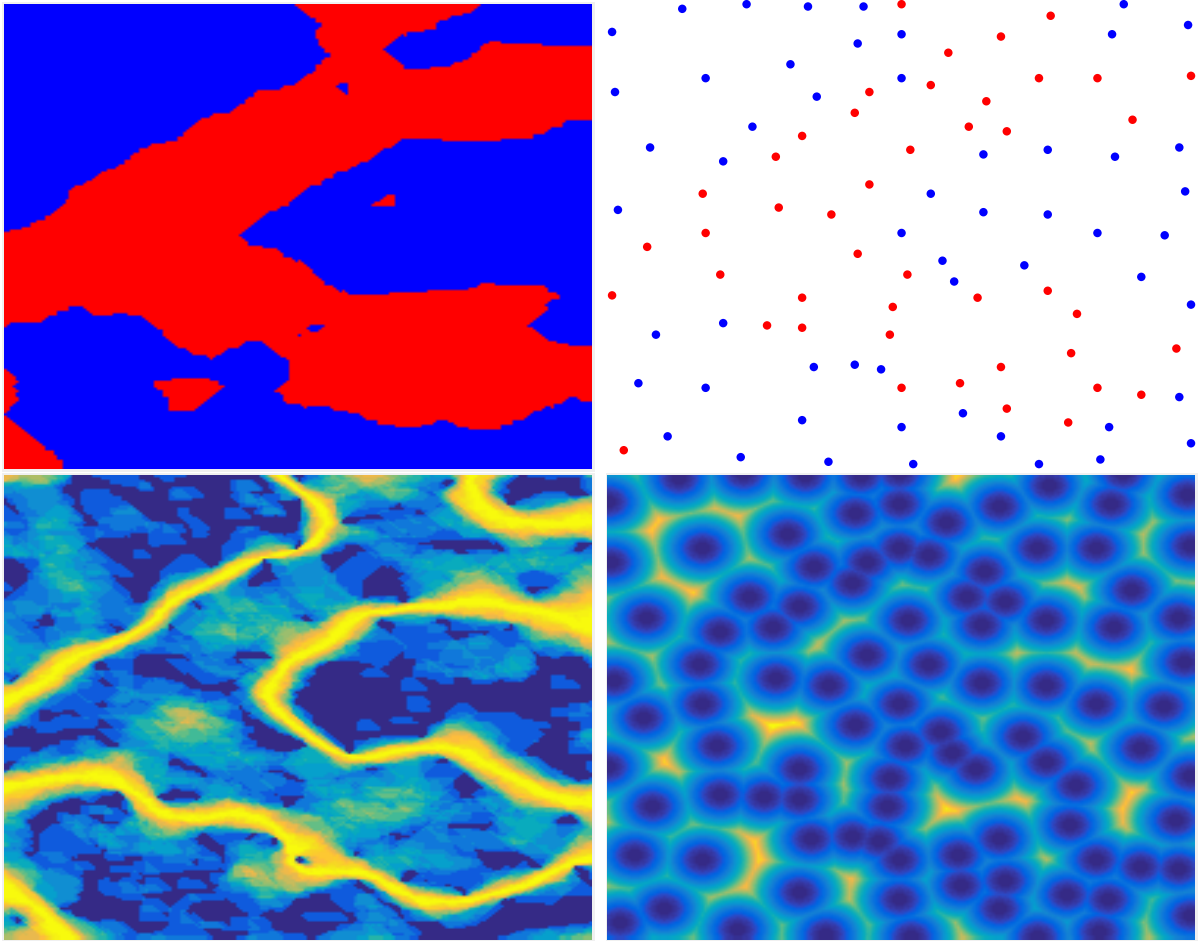


Figure 3.9: Estimated maps for the model SC_1 at $k = 100$. Upper Row, Left: Reference Image, Right: Sampled Locations. Lower Row, Left: Entropy Map ($\alpha = 1$), Right: Distance Map ($\alpha = 0$). Color maps, linear from blue to bright yellow (from low to high entropy or distances)

Figure 3.10 illustrates the resolvability in (3.49) and average frequency of error on detecting non-sensed positions in (3.50) for the RAMIS solutions as a function of α for the case of 500 samples (1.25% of the image size). For the three models, a consistent improvement can be observed for a mixed solution with α in the range $[0.7 - 0.8]$. More generally, as seen in Fig. 3.11, the performance improves by the use of a mixed approach as it was expected from the Markov analysis. While in the regime of very few samples the gain attributed to the mixed solutions is not that significant, it increases as the sample number grows (from 0.5% to 2% samples). Finally, the improvement continues but is less noticeable as the sampling rate increases (see the curves in Fig. 3.11 close to the 10%), which is justified since the missing information of non-sensed locations for both schemes is extremely low.

Looking more carefully at the role of the model complexity in this analysis, a positive correlation between the performance gain of the best proposed scheme with respect to the pure structured sampling ($\alpha = 0$) and the field spatial complexity is observed.

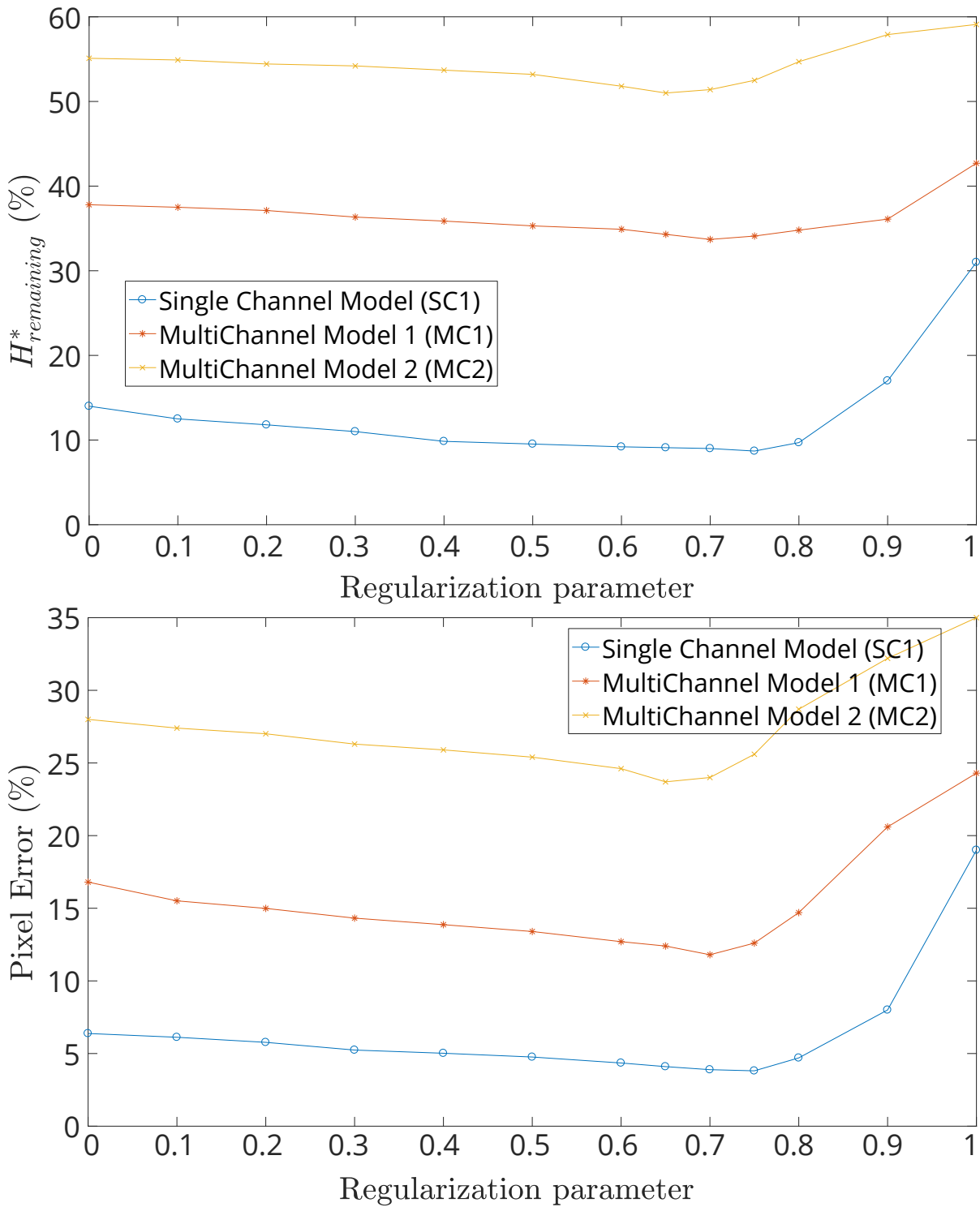


Figure 3.10: Performance of non-sensed positions under RAMIS as a function of α , after 500 samples. Top: Resolvability, Bottom: Mean error. Average curves for 50 independent train-test sets

For example, as summarized in Table 3.1, considering a sampling rate of 1.25% (500 samples), for the three models under analysis the optimal mixing parameter is around

Table 3.1: 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

$\alpha \approx 0.7$. Comparing the mixed solution with respect to the pure regular sampling (reference performance for $\alpha = 0.0$), the proposed solution achieved an overall performance improvement in terms of both absolute gain in resolvability reduction and mean pixel error. Finally, Fig.3.11 presents those gains in performance for the three models and for an extended sampling regime but considering images of size 50 by 50 pixels. Here, in terms of resolvability, $\alpha \approx 0.7$ is consistently a good solution for every model and for a large range of sampling rates.

3.7.6 Comparison with non preferential sampling methods

Finally, the proposed framework is compared (using the regularization parameter $\alpha = 0.7$) with two non-preferential sampling schemes provided by a uniform sampling of the space and the regular grid sampling, which is achieved with the implemented algorithm when $\alpha = 0$. For this analysis, a sampling rate in the realistic range of 0% - 2% of the field is evaluated. For the performance analysis, the resolvability is considered, that is, the pixel average conditional entropy of non-sensed positions given the sensed position values in (3.49), and the frequency error of the simulation of non-sensed positions given in (3.50). For the computation (estimation) of these two indicators, 600 conditional MPS simulations are used to estimate conditional entropies and the frequency of errors. Finally, in order to obtain representative results for the three models used, the average performance across 50 scenarios is studied, where in each scenario a different training and a test image are selected from a collection of unconditional simulations.

Equipped with these two indicators, a global and a local analysis of the results are conducted. For the global analysis, the average indicator (conditional entropies and simulation error) for the complete set of non-sensed positions is computed, and for the local analysis, the focus is on the transition zones between low and high values of the true image. This local analysis seems more relevant from the point of view of resolving the categorical nature of channelized fields, because transition zones define the problem. In fact any adaptive sampling method needs to address a compromise between resolving locally or resolving globally the field and, consequently, it is insightful to consider these two regimes of analysis. For the local indicator, the true image is used to define zones around the transitions to compute the average indicator. In particular, all pixels at a distance no

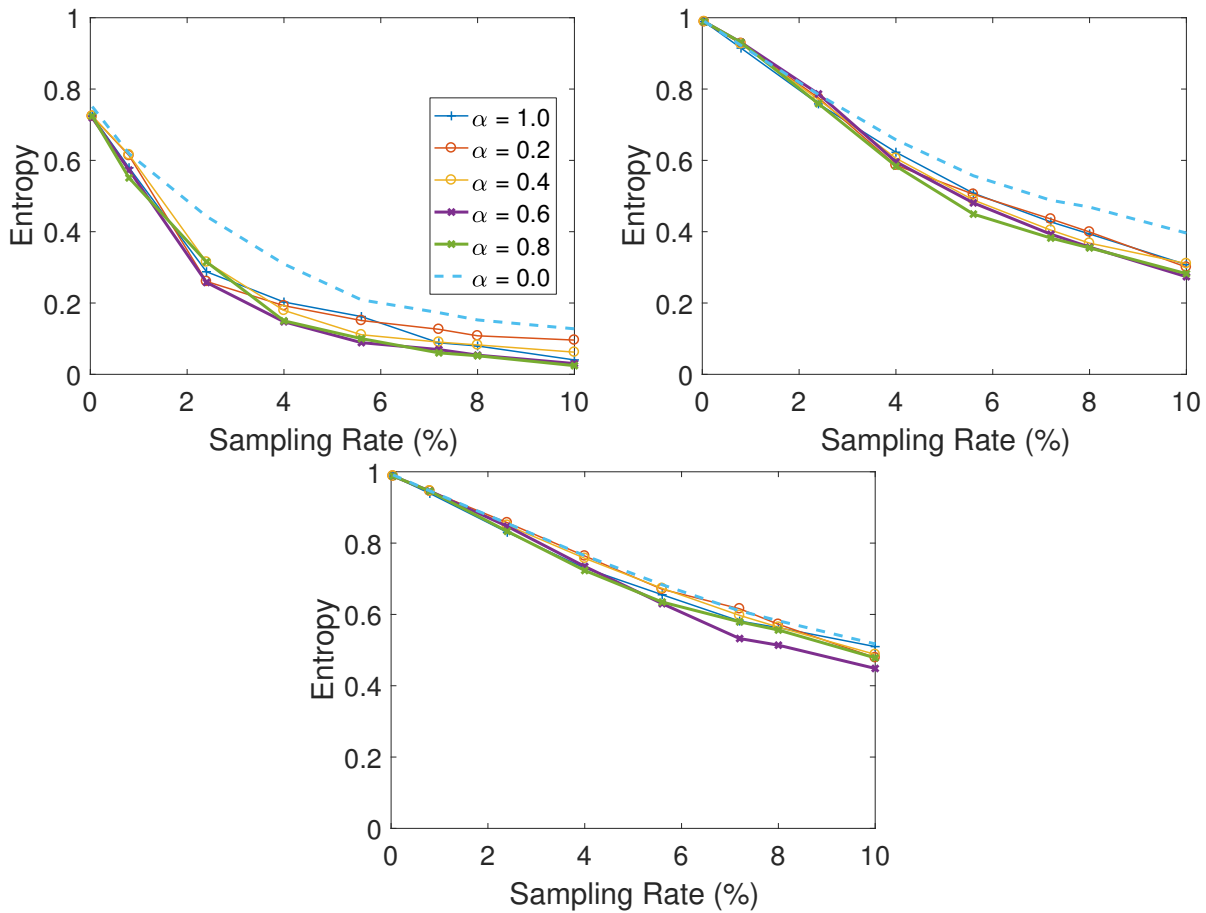


Figure 3.11: Resolvability of non-sensed positions using RAMIS approach as a function of sampling rate for different values of α . (Top left) SC_1 , (Top right) MC_1 , (Bottom) MC_2

higher than 5 pixels from the line of transitions are considered to define the local zones, as can be seen in Fig. 3.12.

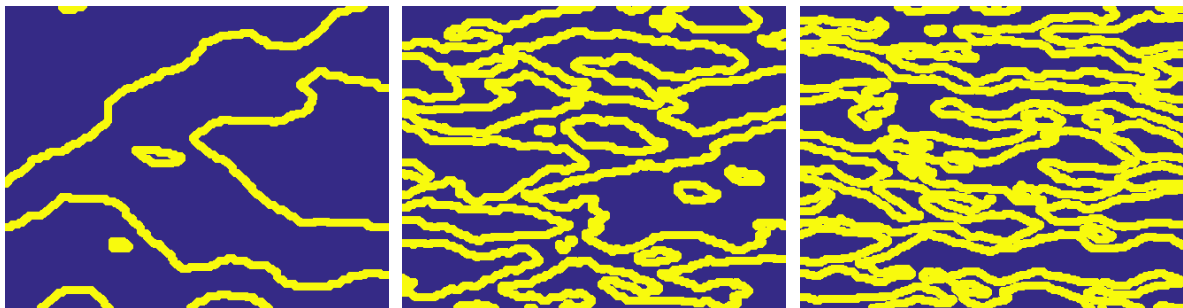


Figure 3.12: Example of the masks used to define transitions in channelized images. 5 pixels around the transitions are considered. From left to right: models SC_1 , MC_1 , and MC_2 . Color maps, solid yellow is the mask

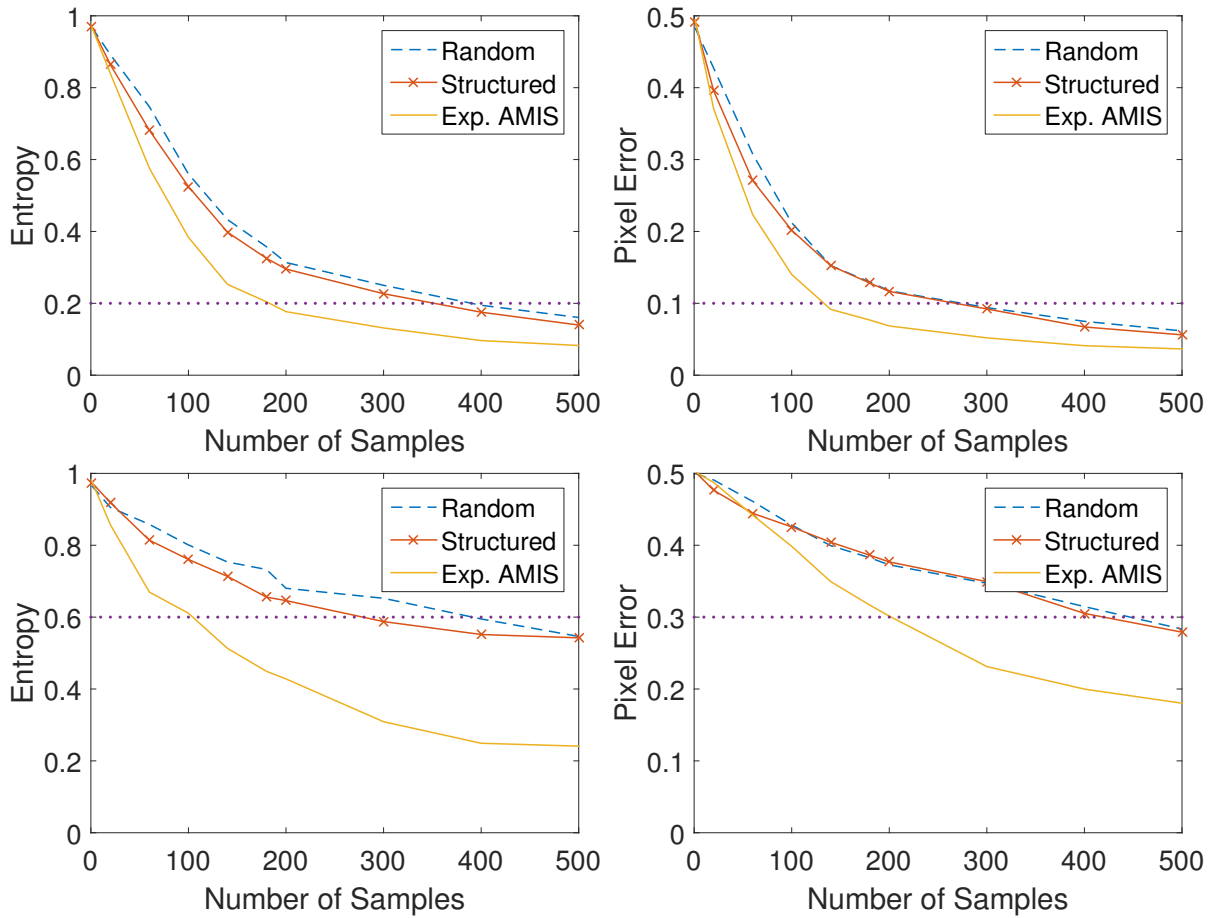


Figure 3.13: Experimental performance for model SC_1 in terms of (left) Remaining Entropy reduction and (right) Estimation error. Top: global performance. Bottom: local performance. Sampling rates from $[0\% - 1.25\%]$ with $\alpha = 0.7$. Mean behavior over 50 realizations

Analysis of the Single-channel Model

Starting with the single-channel model, Fig. 3.13 shows the trend of the average performance (globally and locally) of the studied sampling approaches with the adopted metrics as a function of the sampling rate from 0% to 1.25%.

Remarkably for all the sampling regimes, the improvements obtained by the mixed adaptive approach in terms of resolvability and frequency of error both globally and locally are significant and systematic. It is important to highlight that the local gain is substantial, meaning that the information driven adaptive strategy helps to better characterize transition zones than a pure random or a structural sampling. The local gain is very significant in the regime of $0\% - 1.25\%$ sampling rate for both metrics. This important gain in local characterization does not imply detrimental performance at the global level, which is relevant because this implies that the price of putting samples locally around transitions does not degrade the resolvability and estimation of zones far from the transitions. In conclusion, for the single channel model, a significant gain is observed in

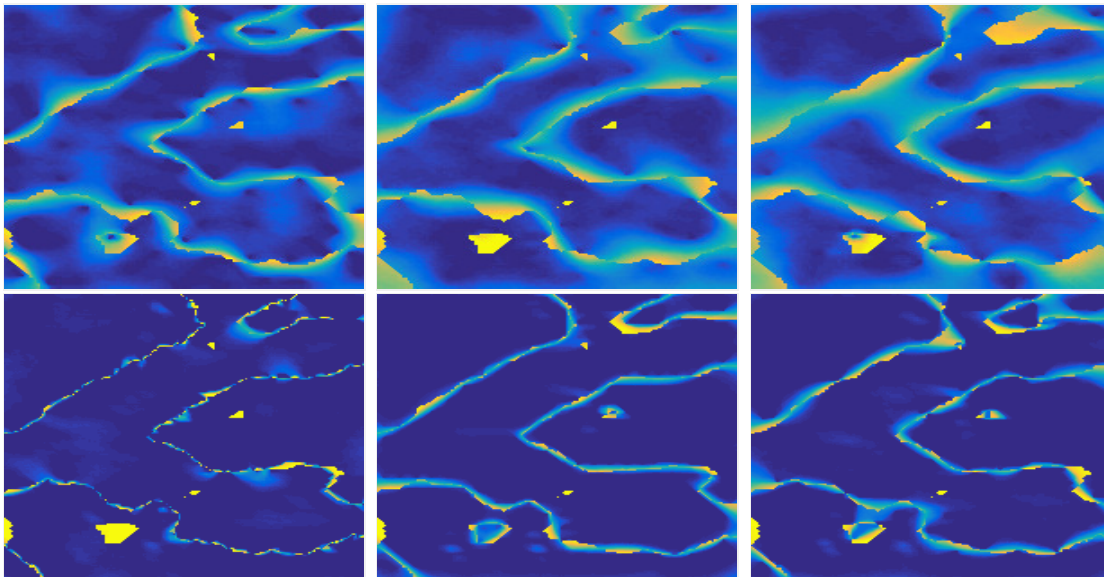


Figure 3.14: Remaining Entropy maps for model SC_1 using 600 realizations of the sampling process. Top: maps using the first 100 samples. Bottom: maps using the first 500 samples. Left to right: RAMIS, quasi-regular, and random sampling. Color maps, linear from blue (low remaining entropy) to bright yellow (high remaining entropy)

the ability to recover the true image from the adoption of the proposed adaptive scheme.

To illustrate these improvements at the global level, it can be observed that the non preferential sampling schemes require a sampling rate of 0.875% (350 samples) to obtain a conditional entropy of 0.2 (see Fig. 3.13), while the proposed approach only requires a sampling rate of 0.45% (180 samples) to achieve the same objective. A similar trend is observed for the simulation error, where the proposed method requires a sampling rate of 0.375% (150 samples) to achieve a simulation error of 0.1 while the classical approaches require a rate in the order of 0.75% (300 samples). Therefore, the adaptive strategy offers a reduction of 50% on the number of samples to achieve similar results than with non preferential sampling. If considering the local performance, the reduction in sampling rate is even more dramatic.

Finally, Fig. 3.14 shows a map of the conditional entropies for the three sampling methods at the sampling rates of 0.25% and 1.25%, respectively, showing the effectiveness of the proposed scheme as it presents better resolvability (smaller conditional entropies) on the transition zones.

Analysis of the Multi-channel Models

For the case of multi-channel models, the results are reported in Figs. 3.15 and 3.16. First, it is important to comment on the dependency of these results on the model complexity. As expected the number of samples required to achieve the same reconstruction error and resolvability increases with the complexity of the model. For example, for the single

channel model the proposed method can achieve a resolvability of 0.2 with a sampling rate of only 0.45% (180 samples). On the multi channel models, the proposed method requires sampling rates over 2% (800 samples) (see Figs. 3.17 and 3.18). For this reason, in all these experiments a higher number of samples, from 1 – 800, is considered.

In general, the RAMIS approach shows consistently better performances in resolvability and simulation errors than the conventional non-adaptive scheme globally and locally. As it was shown before for the single-channel model, this performance improvement is more important locally near transition zones, where the inference task is known to be more challenging, than far from the transition zones. However the regime of gains and their magnitude are reduced compared with the results obtained for the single-channel model. The reason could be that for higher complexity models a much higher sampling rate is needed to move the problem to a regime where adaptive samples make a significant contribution in the characterization of the field. In particular, there is a range from 0–100 samples for the first multi-channel model where no major differences in performances are observed globally between the implemented method and the non-adaptive schemes (see Fig. (3.15)). This regime increases to 0 – 200 samples for the second multi-channel model in Fig.3.16. Despite this observation, the local gains are relevant and significant and there is a regime of higher sampling rate where the RAMIS method shows a gain in resolvability and in the rate of detecting non-sensed positions from the simulations.

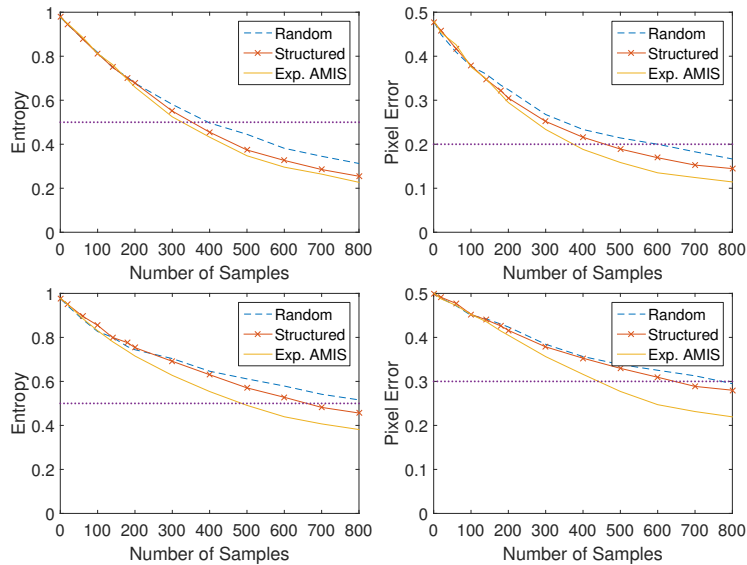


Figure 3.15: Experimental performance for model MC_1 . Left: Remaining Entropy reduction, Right: Estimation error. Top: global performance. Bottom: local performance. Sampling rates from [0% – 2%] with $\alpha = 0.7$

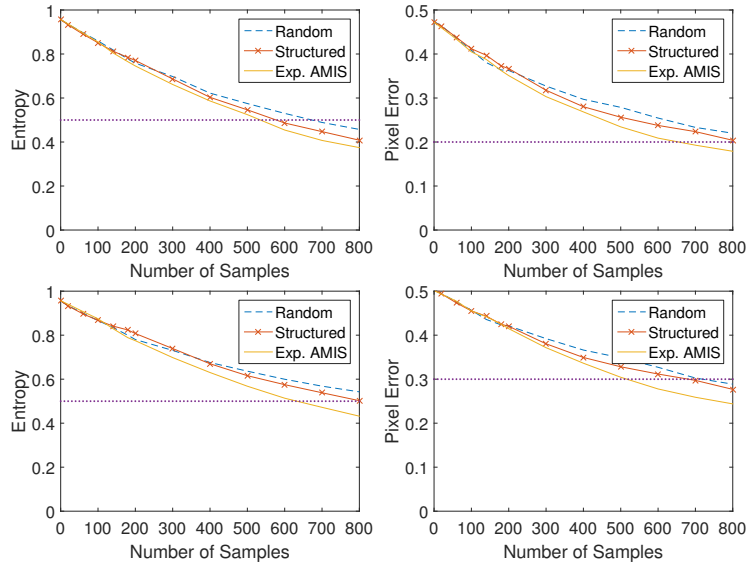


Figure 3.16: Experimental performance for the model MC_2 . Left: Remaining Entropy reduction, Right: Estimation error. Top: global performance. Bottom: local performance. Sampling rates from $[0\% - 2\%]$ with $\alpha = 0.7$

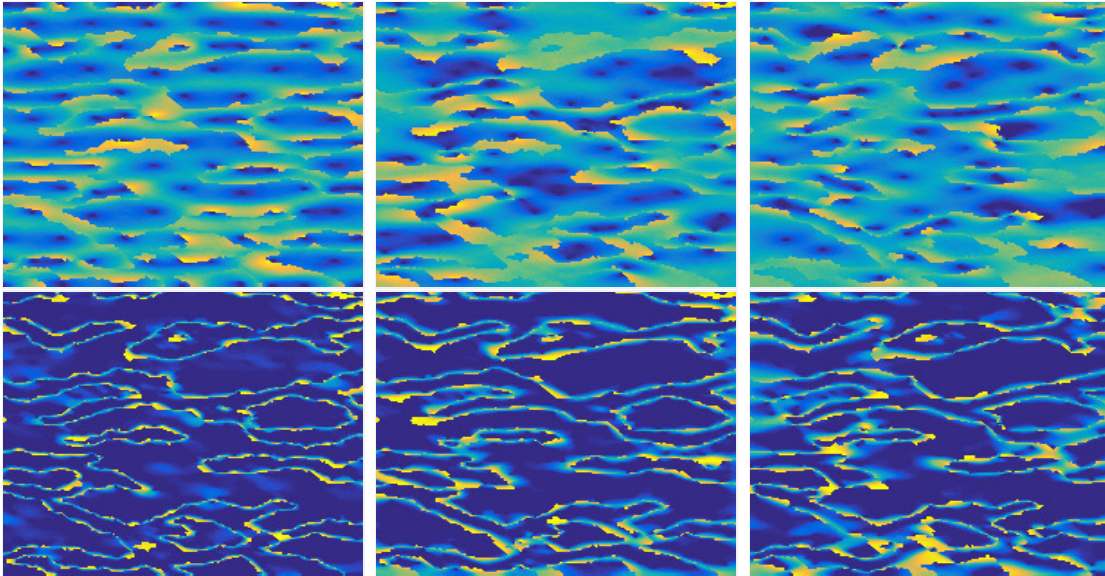


Figure 3.17: Remaining Entropy maps from 600 realizations for model MC_1 . Top: using only the first 100 samples. Bottom: for the first 800 samples. Left to right: RAMIS, quasi-regular, and random sampling. Colormap: same as fig 3.14

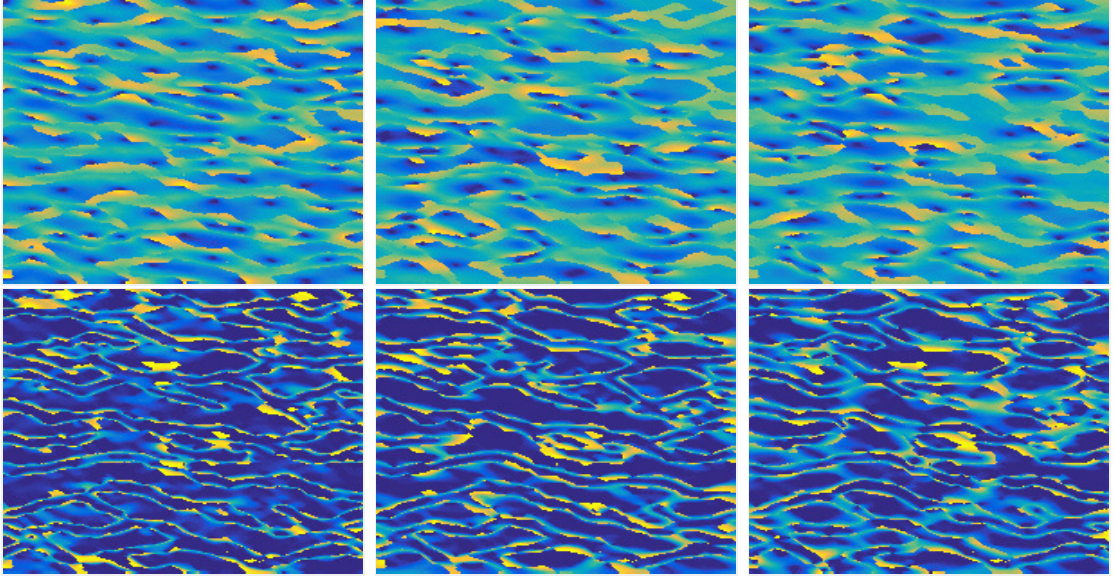


Figure 3.18: Remaining Entropy maps from 600 realizations for model MC_2 . Top: using only the first 100 samples. Bottom: using the first 800 samples. Left to right: RAMIS, quasi-regular, and random sampling. Colormap: same as fig 3.14

3.8 Discussion and Final Remarks

In this chapter, the role of preferential sampling is explored for the task of multiple-point simulation (MPS). For this, the problem of optimal sampling is formalized and addressed from a maximum information extraction perspective. This sampling principle has the ability to locate the samples adaptively on the positions that extract maximum information for the task of characterizing the underlying field. A formal justification is provided for adopting this information-driven sampling criterion and show its practical benefits for MPS in the context of simulating channelized facies models. Interestingly, this preferential sampling locates samples adaptively on the transition between facies and this improves the performance of conventional MPS algorithms and, consequently, it can be integrated into the framework of MPS. In conclusion, preferential sampling can contribute in MPS even at very small sampling regimes, and as a corollary, the training image can be used constructively not only to simulate non-sensed positions, but also to decide where to measure next.

The Appendix A provides details of the implementation and formalization of the proposed methods presented in this chapter.

Chapter 4

Improving Short-Term Planning and Grade Control with an Entropy-based Adaptive Sampling Strategy

Grade control and short-term planning determine the performance of a mining project. Improving this decision, by collecting the most informative samples (data) may have significant financial impact on the project. In this research, a method to select sampling locations is proposed in an advanced drilling grid for short-term planning and grade control in order to improve the correct assessment (ore-waste discrimination) of blocks. The proposed sampling strategy is based on a regularized maximization of the conditional entropy of the field, functional that formally combines global characterization of the field with the principle of maximizing information extraction for ore-waste discrimination¹. This sampling strategy is applied to three real cases, where dense blast-hole data is available for validation from several benches. Remarkably, results show relevant and systematic improvement with respect to the standard regular grid strategy, where for deeper benches in the field the gains in ore-waste discrimination are more prominent.

Short-term planning and grade control aim at determining the optimum assignment of each block of material in a mine, considering the potential economic profit, complying with the mine plan and the constraints in the mine and processing facilities. This assignment can be a specific process, a stockpile or the waste dump. In this chapter, the analysis has been limited to the binary decision case of ore sent to the processing plant or waste sent to the dump. In this context, the grade and other properties of each block are estimated from samples located in its neighborhood. In most open pit mines, these samples are taken where blast-holes are drilled. Two important considerations about this estimation procedure are:

¹The content of this chapter is based on the **Santibañez-Leal † F**, Silva JF, Ortiz JM (2019) Improving short term planning and grade control with an entropy-based adaptive sampling strategy. Natural Resources Research online first:1–35. Submitted. Under Revision

- The block grade estimation involves a change of support, that is the volumetric support of the samples is usually much smaller than that of the block; and
- The block grade is estimated based on the most current information, which during production is the sampled grades obtained at blast-holes.

Because of these considerations, the estimation technique of the block grade must account for the *support effect* as the block ordinary kriging is typically used to determine the block grade. Secondly, the fact that the assignment of the block to a specific destination is based on an estimation from limited data, means that sampling errors in that information impair the final assignment (Chiles and Delfiner, 2012, p. 7) (Abzalov et al, 2010), which is known as *information effect* (Chiles and Delfiner, 2012, p. 442)

Short-term planning and grade control are almost always based on samples taken from a pseudo-regular grid of blast-holes. These samples suffer from high errors due to numerous factors: poor sample recovery near the collar of the blast-hole, delimitation error by including the subdrill, and careless applications of sampling procedures since production has priority and the area must be freed prior to loading explosives (Francois-Bongarcon, 1983; Ortiz et al, 2012; Pitard, 1993). Poor sample quality leads to extremely large economic losses, which increase when geological domains are poorly understood or when estimation of the block grades suffers from large errors (Magri and Ortiz, 2000). These sources of uncertainty are hidden, but have a significant impact in the financial performance of mining projects (Magri et al, 2010, 2011; Ortiz and Deutsch, 2004).

Regarding the information used for grade control, most applications in mining deal with a regular grid, and optimization is aimed at determining the spacing of the samples to comply with some measures of quality (Ortiz and Magri, 2014). These measures of quality are often linked to the kriging variance, as the measure of performance of the estimation (Blackwell, 1998; Hassanipak and Sharafodin, 2004; Lloyd and Atkinson, 1998; McBratney and Webster, 1981; McBratney et al, 1981; Vasát et al, 2010). The use of geostatistical simulation allows to consider uncertainty that may be a function of the block grade, accounting for the *proportional effect* (Journel and Kyriakidis, 2004).

Furthermore, although in most applications conventional two-point simulation methods, such as sequential Gaussian or sequential indicator simulation are used, some authors have introduced the use of multiple-point geostatistical simulation in mining applications (Boisvert et al, 2008; Ortiz, 2003; Ortiz and Deutsch, 2004). These methods require the use of a training image (Mariethoz and Caers, 2014), which is commonly built from an analogue, outcrop or a geological interpretation (Boisvert et al, 2008). Data driven training images can also be used (Ortiz and Deutsch, 2004; Silva Maureira, 2015) and are precisely the approach taken in this study.

Advanced drilling considering a reverse circulation (RC) drilling rig has been studied as an option to improve the quality of the samples for short-term planning and grade control, and to provide grades and other geological information in advance to build a reliable model for decision-making. This approach has led to improved financial returns (Magri et al, 2010, 2011; Ortiz and Magri, 2014; Ortiz et al, 2012) and allows the use of more sophisticated tools to build the short-term plan, such as cokriging or cosimulation to

account for multivariate relationships. Furthermore, advanced drilling allows a characterization of the geological features of the rock (lithology, alteration, mineralization), hence these samples can be used to improve and update the geological interpretation. Advanced drilling is normally applied considering a regular drilling grid with a spacing wider than the blast-hole spacing. From the high quality samples obtained in the advanced drilling grid, a short-term model can be built using geostatistical estimation (kriging or cokriging) or simulation (Journel, 1974). The simulation approach has usually relied on a multiGaussian method and to the best knowledge of the authors, multiple-point geostatistics has not been applied in grade control. Cost functions that account differently for the costs of misclassification as ore or waste have also been used frequently in mining (Deutsch et al, 2000; Dimitrakopoulos and Godoy, 2014; Journel and Kyriakidis, 2004; Verly, 2005).

This work presents a new methodology for short-term planning sampling and grade control based on the selection of the sampling locations that are the most informative in terms of ore-waste discrimination. The proposal aims at characterizing the contact between ore and waste, by learning from the data of previously mined benches the location and geometry of the contact, and determining the best locations to be sampled at the current bench. The implementation of this principle results in an irregular grid of samples. Furthermore, the sample locations adapt to the local conditions of the problem. New samples are drilled at locations that have the highest conditional entropy based on the available samples (within the bench) and on the spatial continuity of the ore and waste blocks. This last component of the model is cleverly estimated from previously mined benches. On the specifics, an algorithm is adopted to this problem to select the best locations, previously introduced by the authors in **Santibañez † F** et al (2019), (Section 4.1). This methodology is applied to three real scenarios and then the results for each case are analyzed, showing the performance improvement with respect to an advanced regular sampling grid. Then, the limitations and potential economic benefits of this approach are discussed and some final conclusions are provided.

4.1 Entropy-based Adaptive Sampling Strategy

In this section, the proposed method for determining the best sample locations over a binary two-dimensional field is reviewed. This strategy is based on a regularized maximum entropy sampling problem presented in **Santibañez † F** et al (2019), which provides a sampling scheme that maximizes the information extracted from the measurements.

4.1.1 Maximum Entropy Sampling

To formalize the problem, let consider a $2-D$ field with unknown spatial correlation. Notice that this spatial correlation may be fully characterized by a variogram or may require higher order statistics to be described (Mariethoz and Caers, 2014; Ortiz, 2003). Then, and without loss of generality, the regionalized variable X is a $2-D$ random array

of variables representing a discrete image of finite size $A \cdot B$,

$$X_{u,v} : (\Omega, \mathbb{P}) \rightarrow \mathcal{A} = \{0, 1\} \quad \forall (u, v) \in [A] \times [B], \quad (4.1)$$

with values in the finite alphabet \mathcal{A} (limited here to the binary case), and (Ω, \mathbb{P}) describing the sample space and the probability measure for the stochastic regionalized variable.

Adopting the concept of entropy as a measure of uncertainty of a random variable (Cover and Thomas, 2006), an algorithm that finds the placement rule f through optimal reduction of the posterior entropy after doing measurements is proposed. For any given number $K \leq A \cdot B$ of sample positions to be taken, let $\mathbf{F}_K \equiv \{f : \{1, \dots, k, \dots, K\} \rightarrow [A] \times [B]\}$ be the collection of functions that select K -elements in $[A] \times [B]$. Then for any $f \in \mathbf{F}_K$ (sampling rule of size K) let's denote the measured random vector by:

$$X_f \equiv (X_{f(1)}, X_{f(2)}, \dots, X_{f(K)}), \quad (4.2)$$

and the remaining non-measured random vector by,

$$X^f \equiv (X_i : i \in [A] \times [B] \setminus f). \quad (4.3)$$

In this context, the conditional posterior entropy of X^f given X_f measures the remaining uncertainty of the field X after sensing the position in X_f (Cover and Thomas, 2006). It can be computed as the joint entropy of the entire process minus the joint entropy of the variables measured by f , as shown in Eq. (4.4).

$$H(X^f | X_f) = H(X) - H(X_f). \quad (4.4)$$

Then, the optimal decision of size K is the solution of

$$f_K^* \equiv \arg \min_{f \in \mathbf{F}_K} H(X^f | X_f), \quad (4.5)$$

which minimizes the uncertainty of the field after K measurements. Interestingly, adopting some information theory identities (Cover and Thomas, 2006), the optimal decision in (4.5) is equivalent to the solution of the *maximum entropy problem* (Santibañez † F et al, 2019):

$$f_K^* = \arg \max_{f \in \mathbf{F}_K} H(X_f) \quad (4.6)$$

that finds the K positions that jointly lead to the highest a priori (before the measurements) entropy. This principle is easier to implement as it requires marginal distributions and not the complete joint distribution of X , model for the implementation of the Eq. (4.5).

The Sequential Adaptive Strategy From the general principle in Eq.(4.6), the focus is on the realistic *sequential problem* where the decision is taken sample by sample, and furthermore, the measurements taken in previous iteration of the algorithm are considered to upgrade the model (or adapt the model to the data) in the next iteration. Considering the set of previous sensed position $f_k^* = ((i_1, j_1), \dots, (i_k, j_k))$ and the measurements of

the field X taken at those selected places, i.e., $\bar{x} = (x_1, \dots, x_k)$, the solution of the $k + 1$ position is given by the maximum entropy principle in (4.6) conditioned on $X_{(i_1, j_1)} = x_1, \dots, X_{(i_{k-1}, j_{k-1})} = x_{k-1}$ and $X_{(i_k, j_k)} = x_k$, i.e.,

$$(\hat{i}_{k+1}, \hat{j}_{k+1}) = \underset{(i,j) \in [A] \times [B] \setminus \{(i,l): l=1, \dots, k\}}{\operatorname{arg\,max}} H(X_{i,j} | X_{f_k^*} = \bar{x}), \quad (4.7)$$

where $f_{k+1}^* = (f_k^*, (\hat{i}_{k+1}, \hat{j}_{k+1})) \in \mathbf{F}_{k+1}$. Then iterating this rule from $k = 1, \dots, K$ offers an adaptive and sequential solution for the problem of selecting the more informative K positions of the field.

The Regularized Adaptive Strategy The adaptive solution in Eq. (4.7) requires the knowledge of the statistics of X . In practice, this model is not available requiring to find a way to infer this model from empirical data. In the proposed solution, a training image is required that is deemed to represent the spatial continuity of the random function describing the spatial correlations in the underlying spatial model, and from this a model is estimated using conditional multiple point simulations (Mariethoz and Caers, 2014; Ortiz and Deutsch, 2004). In particular, a set of conditional probabilities of the form $\hat{\mu}_{X_{i,j}|X_S}(x_{i,j}|x_S)$ are estimated, where $i, j \in [A] \times [B]$, $S \subset [A] \times [B]$ denotes the conditioning positions (attributed to sensed data), and $x_{i,j} \in \mathcal{A}$ and $x_S = (x_{i,j} : (i, j) \in S) \in \mathcal{A}^{|S|}$, which is the information needed to implement (4.7).

Finally, a convex combination between the maximal entropy adaptive criterion in (4.7) and a regularization principle that promotes uniform covering of the sampling space is proposed. It has been found that a mixed rule that promotes a compromise between selecting the most informative points (from the conditional model and previous data) and a good cover of the sampling space is better than the rule based on a pure adaptive principle (Santibañez † F et al, 2019). More precisely, let $S_k = \{(i_l^a, j_l^a) : l = 1, \dots, k\}$ denote the collection of k positions previously obtained. $X_{S_k} = \{X_{i_1^a, j_1^a}, \dots, X_{i_k^a, j_k^a}\}$ denote the random vector at the locations S_k , and $x_{S_k} = \{x_1, \dots, x_k\}$ be the data collected at S_k . Then, the regularized rule is the solution of

$$(\hat{i}_k^a, \hat{j}_k^a) = \underset{(i,j) \in [A] \times [B] \setminus S_{k-1}}{\operatorname{arg\,max}} \alpha \cdot H(X_{i,j} | X_{S_{k-1}} = x_{S_{k-1}}) + (1 - \alpha) \cdot D((i, j), S_{k-1}). \quad (4.8)$$

Note that the second term promotes a uniform sampling by using a distance criterion

$$D((i, j), S_{k-1}) = \min_{(\tilde{i}, \tilde{j}) \in S_{k-1}} d((i, j), (\tilde{i}, \tilde{j})), \quad (4.9)$$

where $d((i, j), (\tilde{i}, \tilde{j})) \equiv \sqrt{(\tilde{i} - i)^2 + (\tilde{j} - j)^2}$. Therefore, the proposed regularized rule corresponds to a global maximin strategy.

4.2 Optimal Sampling for Grade Control

In grade control the target is to define the destination of every single block in the current mining bench. In order to achieve this task, the estimated grade of the block is considered and compared with the *Cut-Off* grade. The *Cut-Off* grade corresponds to the minimum grade required for a block to be considered for processing (expecting a positive economical benefit). Therefore, a block found to be above this *Cut-Off* grade is considered to be ore, while a block below this grade is considered to be waste. Then, the grade control relies on a binary decision even when the value of the block grade corresponds to a continuous variable.

The *Cut-Off* grade can be established by various methods. Its selection is related to a certain production objective, such as the use of resources or economic benefit. These objectives give rise to different types of targets, such as maximizing global economic benefits, immediate benefits and so on. The *Cut-Off* grade does not have a fixed or predefined value, but instead it corresponds to a strategic variable that has important implications in the design and production of the mine.

The application of the adaptive sampling strategy to grade control is described in this section by considering a scenario where several benches are to be mined in a sequential process. The sampling strategy is used with the objective of improving the prediction of ore-waste contact in successive working benches. For every bench, the method works in two stages: first, the general features of the field are learned through a coarse regular sampling grid; then, the sampling adapts to the features learned from the regular grid and aims at characterizing the contacts between ore and waste for a better characterization of non-sensed blocks as learned from the benches previously mined out. In each bench, the process is applied sequentially, where current samples are used to make decisions about the subsequent locations to be sampled.

4.2.1 Proposed Methodology for Sampling

Considering a collection of M benches, where every bench is indexed by $m \in \{1, \dots, M\}$ and it is a rectangular grid, i.e., a 2-D image, denoted by $I_{\text{grade}}^m \in \mathbb{R}^{A \times B}$ of size $A \times B$. On each bench m , K measurements will be taken to infer a model of the mineral distribution of non-sampled blocks at this bench. Therefore, the main problem is to define the best K locations using the framework presented in Section 4.1. From these K samples an estimation is made on all the non-sensed blocks of the bench producing what is defined as the block model. Finally, a hard threshold is applied on the block model using a *Cut-Off* grade to create a binary 2-D image denoted by $I_{o-w}^m \in \mathcal{A}^{A \times B}$ (with $|\mathcal{A}| = 2$), where, consequently, non-sensed blocks are classified as ore or waste.

In the initial stage of this process ($m = 1$), the upper bench is considered. As no preliminary data is available from previous benches, the conditional entropies are drawn from an i.i.d distribution of ore and waste for implementing Eq. (4.8). This worst case

scenario in terms of prior information for the inference, reduces Eq. (4.8) to the classical near regular sampling, since its second term in Eq. (4.9) dominates the optimization promoting distance as a criterion where a uniform coverage of the space is the optimal solution. Thus for the bench $m = 1$, K samples are distributed in a regular grid. An empirical estimation of the remaining (non sensed) mineral grades is performed by 2-D kriging. Finally, after applying the *Cut-Off* grade the image I_{o-w}^1 is obtained.

For benches $m = 2, \dots, M$, the use of the previously estimated binary block model I_{o-w}^{m-1} for bench $m - 1$ is used as the training image (model) for inference of the conditional entropies used in Eq.(4.8) in the current bench. The key assumption made on this selection is that the previous bench $m - 1$ reflects the spatial distribution of the ore and waste blocks more accurately than the i.i.d. assumption made on the initial stage ($m = 1$). Using I_{o-w}^{m-1} as a training image along with *MPS*, in particular the *SNESIM* algorithm, and the previous $k - 1$ measurements at the current bench m , it is possible to estimate the entropy map, $\hat{H}^{m,k}$, for this bench. This is required for the implementation of the sampling rule in Eq.(4.8). The selection rule for the k -th sample at bench m given the previously sampled locations S_{k-1}^m corresponds to the solution of:

$$(i_k^a, j_k^a)^m = \arg \max_{(i,j) \in [M] \times [M] \setminus S_{k-1}^m} \alpha \hat{H}^{m,k}(X_{i,j}^m | X_{S_{k-1}^m}^m = x_{S_{k-1}^m}) + (1 - \alpha) D((i, j), S_{k-1}^m). \quad (4.10)$$

The solution of Eq.(4.10) places the new samples in the locations with an optimal balance between maximum conditional entropy (information criterion) and the maximum distance to previously sampled positions (regularization). It is important to mention that in the proposed practical implementation of Eq.(4.8), the adaptive sequential sampling strategy selects a new batch of s samples at every sampling step. This iterative strategy is repeated until the K samples are obtained. For completeness, a schematic representation of the sampling rule in (4.10) is illustrated in Fig. 4.1. Finally, the pseudo-code of the implementation of Eq.(4.10) is presented in App. B.1.

4.2.2 Classification Process

Concerning the final ore-waste classification, after the K samples are taken in every bench, ordinary kriging is performed using the routine provided in *GSLIB* (Deutsch and Journel, 1998). A variogram model is fit in each case over the full bench, performing the estimation by using a minimum of min_S samples and a maximum of max_S , with a search radius of rad_S [m]. A block discretization of $nA \times nB \times nZ$ points per block is used. The resulting estimated block model for the bench is binarized by applying a *Cut-Off* grade. The resulting binary model (image I_{o-w}^m) represents the estimated ore and waste blocks that are used in short-term planning.

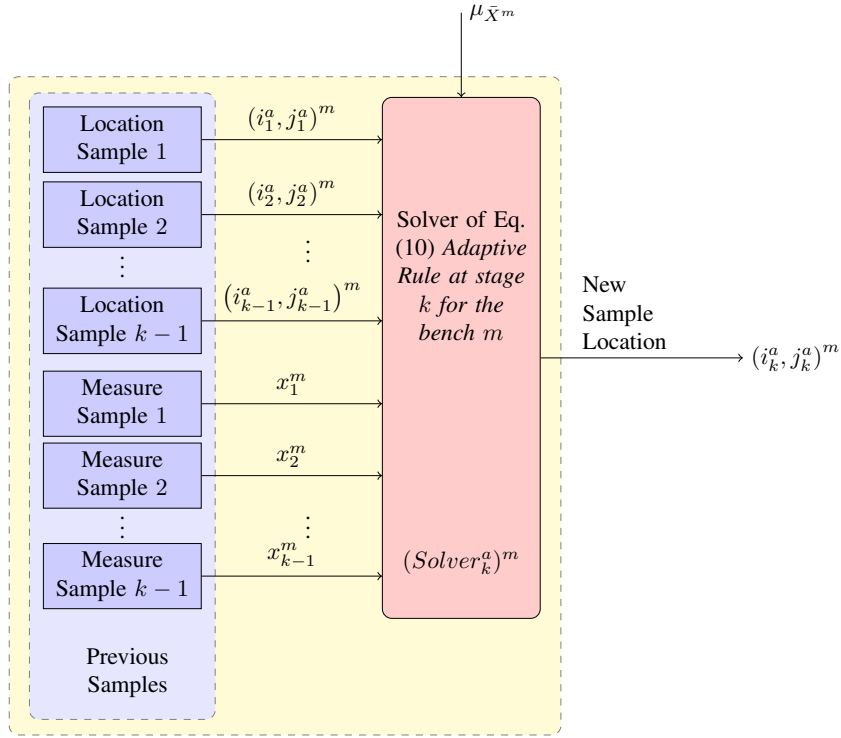


Figure 4.1: Schematic diagram of the sampling rule (4.10).

4.3 Case Studies and Experiments

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 (more details can be found in Section 4.3.1).

In this section, the implementation parameters are provided and the results of the first case study are described in detail. For sake of space, only the relevant results for the other two cases are presented.

4.3.1 Database Description

Case studies 1 and 2 come from the same mining project, but from different areas of the open pit. These areas have already been mined out, providing blast-hole samples to test the hypothesis that adaptive sampling can achieve better discrimination between ore and waste than classical regular sampling. For this mining project, the available blast-hole database consists of nearly regularly spaced samples taken in a grid of approximately $10m$ by $10m$. The database contains areas without samples, therefore, it was split into two informed sectors to generate case studies 1 and 2. On the other hand, case study 3 comes

Table 4.1: 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

Table 4.2: 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

from a different mining project with similar sampling conditions.

Blast-hole data are migrated to a regular grid creating 6 consecutive benches uniformly sampled for each case study. From the selected X - Y section of the raw information provided by the blast-hole samples, a set of dense benches has been created that completely describe the ore distribution of the 3 - D deposit.

The basic statistical information used to build the case studies is summarized in Table 4.1 and Table 4.2.

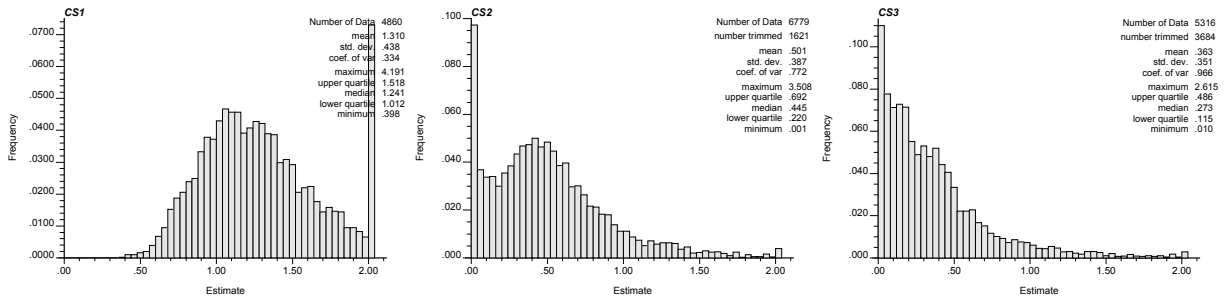


Figure 4.2: Grade mineral distributions and basic statistics for the available blast-holes. From left to right: CS1, CS2, CS3.

Table 4.3: Summary Parameters SNESIM.

	Case Study 1	Case Study 2	Case Study 3
Number of benches (M)	6	6	6
Min data for kriging	4	4	4
Max data for kriging	16	16	16
Max data per octant (0 : not used)	0	0	0
Maximum search radii (x, y, z)	[80,80,10]	[80,80,10]	[80,80,10]
Min data for kriging	10	10	10

The statistical distributions of blast-holes grades present in the analyzed case studies are described in Fig. 4.2, along with their basic statistics.

4.3.2 Construction of Validation Block Model

From the unevenly distributed blast-hole and drill-hole data, a fully informed block model is obtained by performing block ordinary kriging using the kt3d routine of *GSLIB*. Finally, six consecutive benches were considered as ground truth for every case study by considering the blocks inferred from the densely sampled available information. The parameters for the kriging estimation are presented in Table 4.3.

In order to illustrate the density of available information for every single bench, Figs. 4.3 and 4.4 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. 4.5. Systematic descriptions of the other two case studies can be found in Appendix B.2.

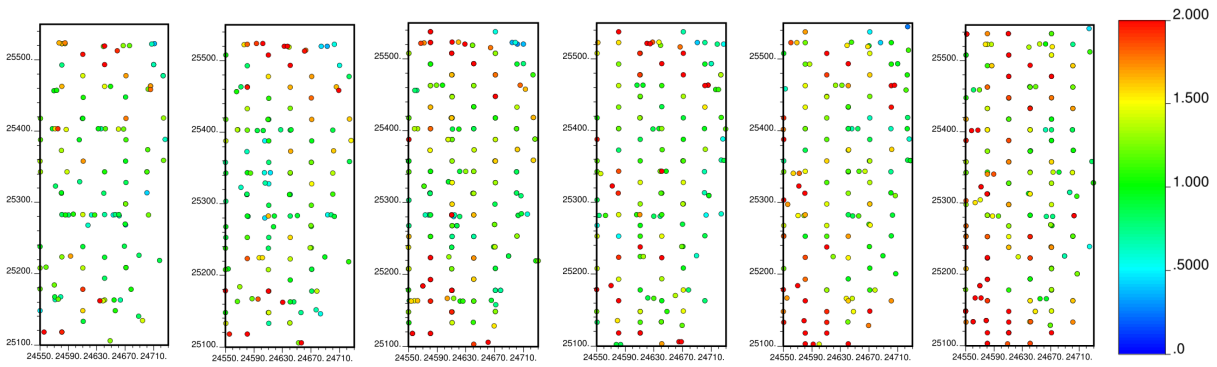


Figure 4.3: Drill-hole samples data for case study 1. From left to right: Benches 1-6. Colormap denotes the grade of Cu .

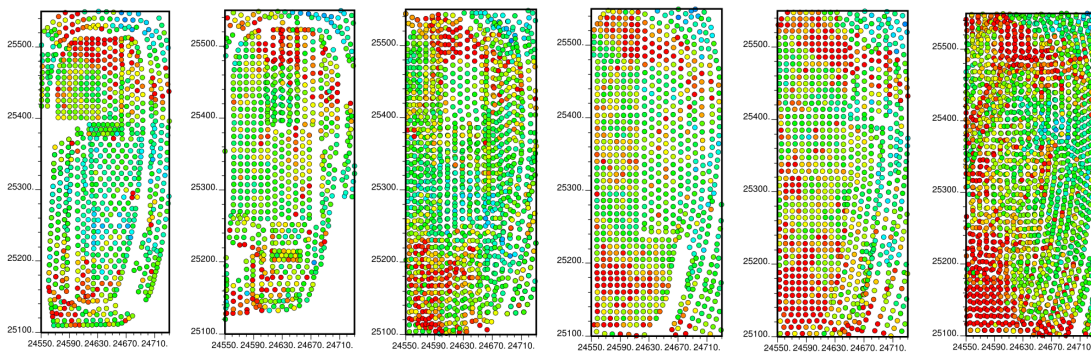


Figure 4.4: Blast-hole data for case study 1. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

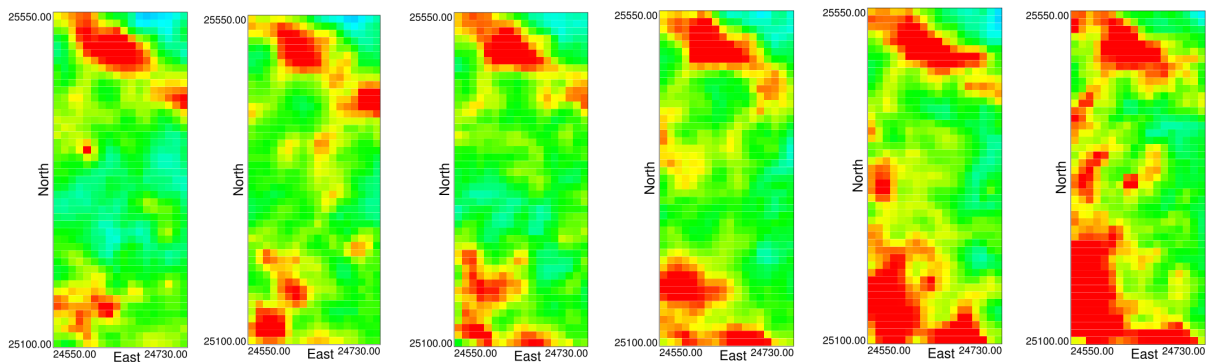


Figure 4.5: Ground truth estimated from drill-holes and blast-holes samples for case study 1. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

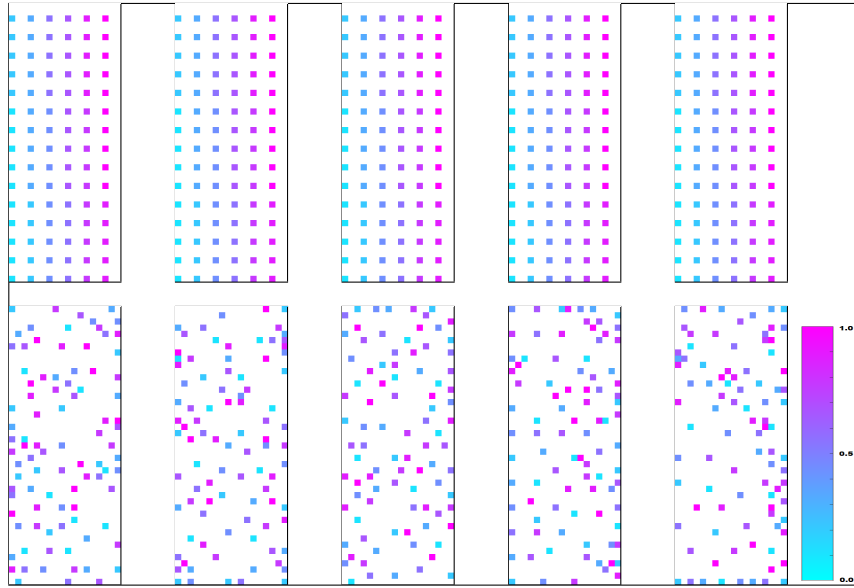


Figure 4.6: Samples for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade 1.012%. Colormap: Describe batch of samples in the order of the performed sampling.

4.3.3 Experimental Results: Adaptive Sampling Strategy

As mentioned, the value for the *Cut-Off* grade corresponds to an operational decision, then in order to demonstrate the benefit of the proposed sampling approach, results for different *Cut-Off* grade values are presented.

Case Study 1

To illustrate the implementation and outcomes of the adaptive sampling strategy, full details are provided for the first case study. For case studies 2 and 3 only summary figures and results are presented, in the understanding that the procedure is similar to that illustrated for case study 1.

For each sampling strategy, 9% of the available locations for each bench were sampled ($K = 0.09 * A_m * B_m$). For case study 1 the proposed samples are presented in Fig. 4.6. The outcomes for kriging estimation from the proposed samples are displayed in Fig. 4.7 for the benches 2, 3, 4, 5 and 6. Three *Cut-Off* grade values, corresponding to the three quartiles of the grade distribution, were considered to evaluate the differences between the classical structured sampling (*STR*) and the proposed adaptive sampling (*ADA*) approach (Fig. 4.8).

From the results presented in Fig. 4.9, it is clear that the number of misclassified blocks is reduced with the adaptive sample in comparison with the structured classical approach.

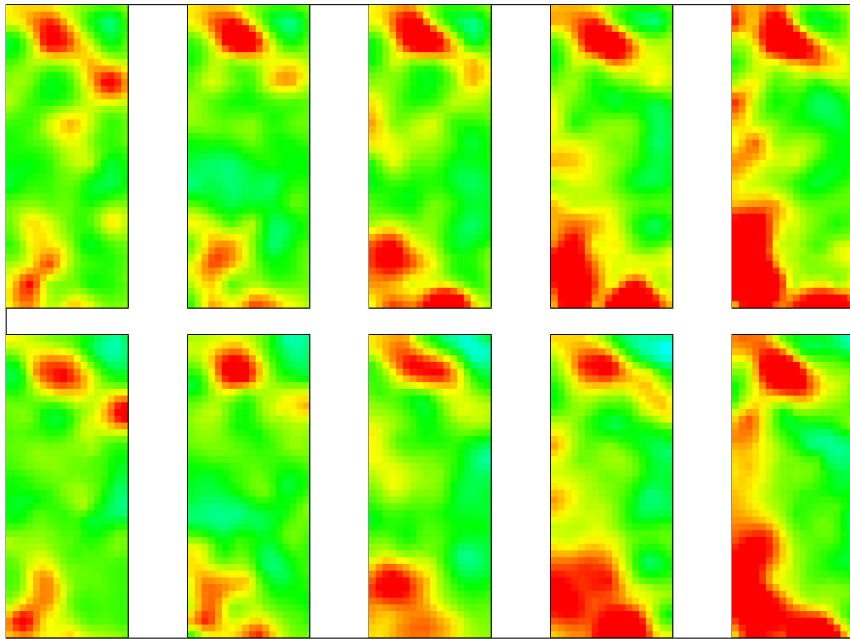


Figure 4.7: Estimated grade for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade 1.012%. Colormap: the same as in Fig. 4.3.

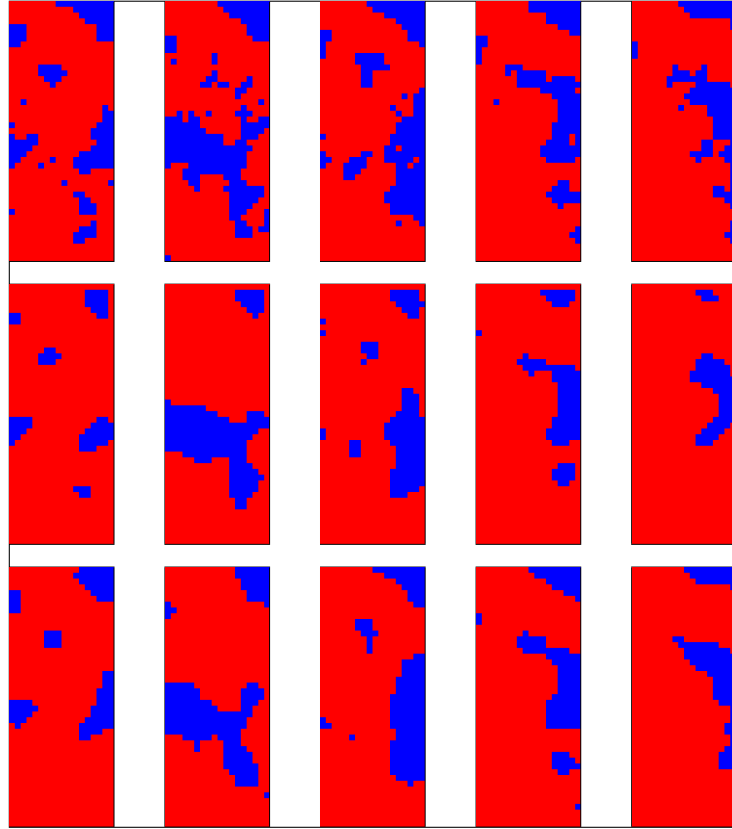


Figure 4.8: Estimated grade control for Case Study 1. From left to right: Benches 2-6. From top to bottom: Ground truth, structured sampling, adaptive sampling using *Cut-Off* grade 1.012%.

		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
True class	Ore	98.9%	1.1%	95.6%	4.4%	98.4%	1.6%	99.1%	0.9%	98.2%	1.8%
	Waste	55.6%	44.4%	36.0%	64.0%	38.8%	61.2%	39.0%	61.0%	55.3%	44.7%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	
		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
True class	Ore	99.2%	0.8%	97.2%	2.8%	97.4%	2.6%	98.9%	1.1%	98.7%	1.3%
	Waste	33.8%	66.2%	28.9%	71.1%	25.4%	74.6%	20.1%	79.9%	28.4%	71.6%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

Figure 4.9: 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%.

Table 4.4: Performance Error Summary for Case Study 1.

	Case Study 1					
	Cutoff grade 1.102		Cutoff grade 1.241		Cutoff grade 1.518	
	<i>STR</i>	ADA	<i>STR</i>	ADA	<i>STR</i>	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

Table 4.5: Performance Error Summary for Case Study 2.

	Case Study 2					
	Cutoff grade 0.220		Cutoff grade 0.445		Cutoff grade 0.692	
	<i>STR</i>	ADA	<i>STR</i>	ADA	<i>STR</i>	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

4.3.4 Performance Assessment

Binary Image Inference Performance We begin by comparing the performances in terms of image recovery achieved by the proposed adaptive sampling strategy with respect to the classical structured sampling. The results are summarized in the Tables 4.4, 4.5 and 4.6 for case studies 1, 2, and 3, respectively. Here the performance is evaluated in the binary ore-waste estimated image for each bench by computing the proportion of misclassified blocks.

From the point of view of the binary allocation of blocks (as ore or waste), the adaptive

Table 4.6: Performance Error Summary for Case Study 3.

	Case Study 3					
	Cutoff grade 0.115		Cutoff grade 0.273		Cutoff grade 0.486	
	<i>STR</i>	ADA	<i>STR</i>	ADA	<i>STR</i>	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

Table 4.7: 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	
	<i>STR</i>	ADA	<i>STR</i>	ADA	<i>STR</i>	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

Table 4.8: 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	
	<i>STR</i>	ADA	<i>STR</i>	ADA	<i>STR</i>	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

sampling based method provides a better overall classification of the blocks. This behavior is consistent for the proposed cut-off grades along all three case studies presented.

In general, as the procedure advances down the benches (as m increases), then the improvement increases as well in terms of the reduction of error as compared to the classical sampling method. Thus, the adaptive strategies outperform the classical sampling method, achieving a consistent improvement over all benches.

It is worth mentioning that it is possible to improve the performance of the proposed technique by feeding the model for the training image of bench m with all the information available from the previous benches ($1, \dots, m - 1$) instead of just from the bench above (bench $m - 1$).

Economic Performance In order to provide a summary of the economic impact related with the sampling strategy in the ore/waste selection process, a brief evaluation of profit is presented taking into consideration some relevant scenarios. The variables and estimations considered for this economic analysis are detailed in Appendix B.4. The achieved results are summarized in Tables 4.7, 4.8, and 4.9.

Table 4.9: 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	
	<i>STR</i>	ADA	<i>STR</i>	ADA	<i>STR</i>	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

Although in some benches the economic results are variable, the overall result obtained shows a systematic improvement over the classic sampling scheme.

Even though the binary assignment as ore/waste is systematically better with the proposed adaptive approach (*ADA*) in terms of mean global error, the current grades are not considered for this evaluation. Therefore, cases can be found where when block grades are close to the *Cut-Off* value, then the classification may fail and the economic value for the bench may decrease.

Negative profits may occur when the processing plants lack ore, and marginal material must be sent to fulfill the production requirements, which may have a negative value, but will likely be higher than dumping these blocks in the waste dump. Now, in most practical cases the *Cut-Off* will be defined to generate profit, rather than minimize the loss.

4.4 Chapter Conclusions

In this chapter, the problem of optimal sampling in the context of short-term planning and the task of classifying blocks to be processed as waste or ore has been addressed. The problem has been formalized and its validation has been presented through the use of subsets of actual mining data.

The proposed methodology takes advantage of the information available from the locations previously sampled, allowing to improve the performance as compared with the classical non-adaptive sampling schemes that has been used for advanced drilling tasks. The proposed strategy has been validated with real blasting data from the exploitation of two copper mines.

From the results obtained across the three analyzed scenarios, it is possible to see that in terms of both error in image reconstruction and global economic value, the proposed methodology achieves better performance than the regular grid based sampling strategy.

The Appendix B provides details of the implementation and formalization of the proposed methods presented in this chapter.

Chapter 5

Conclusions

In this thesis 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 in this thesis for adopting this information-driven sampling criterion as well as concrete ways of implementing this principle in practice. In addition, the practical benefits for *MPS* in the context of simulating channelized facies models is demonstrated using synthetic data and real geological facies. Importantly, this strategy locates samples adaptively on the transition between facies which improves the performance of conventional *MPS* algorithms. In conclusion, this work shows that preferential sampling can contribute in *MPS* even at very small sampling regimes and, as a corollary, demonstrates that prior models (obtained from a training image) can be used effectively not only to simulate non-sensed variables of the field, but to decide where to measure next.

Furthermore, 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 problem has been formalized using the principle of maximum information extraction criterion and the obtained solutions was validated using three data sets of real mining projects. Importantly, 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. From the results obtained across these three real scenarios explored in this thesis, it is possible to see that the proposed methodology achieves better performances than sampling in a structured regular grid (used as a conventional rule for sampling) in terms of both error in image reconstruction and global economic value, when considering the economic revenue of processing the ore and dumping the waste.

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*. Overall the main contribution of this thesis relies on the formulation of the sensing design problem and, as a result of that, the adoption of an adaptive sensing strategy for the characterization of a geological phenomenon. One of the salient aspect of this solution is that the proposed sampling strategy actively uses *MPS* simulations for the estimation of the field statistics. At a higher level, this work confirms the fact that *MPS* simulations can be used to estimate empirical representations of the statistics of regionalized variables for better inference and decision.

5.1 Future Work

Although the presented work was focused on 2-D binary channelized structures (geological facies), the applied principles are general and it 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. Another direction of future work is to study alternative geostatistical simulation tools that could provide more effective estimations of the multi-point patterns (for example using direct sampling). In this direction, the option of estimating the statistics of the model directly from the training image (performing a refined pattern search instead of simulating data) is a very promising.

Finally, although the developed concepts, ideas and algorithms have been developed for inverse problems in geostatistics, 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.

Part of the material developed in this thesis has been left out of the main document for reasons of space. Many of these results were significant and therefore the final appendix of this document is dedicated to cover some of these results.

Appendix A

Complementary Material for Chapter 3

The following sections provide details of the implementation and formalization of the proposed methods presented in the chapter 3.

A.1 App. I: Entropy as an Indicator of Simulation Complexity

Here, a formal interpretation of the *Shannon entropy* as an indicator of complexity is elaborated when the task is to simulate or create a collection of independent and identically distributed (i.i.d.) realizations of a finite alphabet random variable. For that, a connection between simulation and the task of almost lossless source coding (Cover and Thomas, 2006, Chaps. 3.2 & 3.3) will be presented, from which the entropy is a fundamental performance indicator (Cover and Thomas, 2006; Shannon, 1948).

A.1.1 The Operational Complexity Indicator for Simulation

Considering a finite alphabet random variable X taking values in \mathcal{A} with probability $\mu_X \in \mathcal{P}(\mathcal{A})$, the simulation problem of length $n \geq 1$ corresponds to the generation of n -samples in \mathcal{A}^n from the product (or n -fold) distribution $\mu_X^n \equiv \mu_X \times \mu_X \times \cdots \mu_X \in \mathcal{P}(\mathcal{A}^n)$.

A covering argument is considered to stipulate the complexity of simulating i.i.d. realizations of μ_X^n in the product space \mathcal{A}^n . More precisely, *the cardinality of the smallest subset* of sequences in \mathcal{A}^n that captures almost all the probability with respect to μ_X^n , is proposed as a natural indicator of the complexity of simulating μ_X^n in \mathcal{A}^n . This notion can be captured with the following set of definitions:

Definition 1 For $\varepsilon > 0$, $B \subset \mathcal{A}^n$ is ε -typical for μ_X^n , if $\mu_X^n(B) \geq 1 - \varepsilon$.

Definition 2 For $\varepsilon > 0$, the size k is said to be ε -feasible for μ_X^n , if there is $B \subset \mathcal{A}^n$ that is ε -typical for μ_X^n and $|B| \leq k$.

Definition 3 Finally, considering n i.i.d. samples, an indicator of the complexity of μ_x is given by

$$k(\varepsilon, \mu_x^n) \equiv \min \{k : k \text{ is } \varepsilon\text{-feasible for } \mu_X^n\}. \quad (\text{A.1})$$

Adopting these concepts, in particular $k(\varepsilon, \mu_x^n)$, a scenario where a higher number of sequences is needed to capture almost all the probability is more complex (from the point of view of creating i.i.d. simulations), than a case where fewer sequences are needed for the same covering objective. This idea matches the notion of typical set proposed by Shannon to prove source coding theorems (Shannon, 1948). In this context, an interesting aspect to pay attention is the exponential growth of $k(\varepsilon, \mu_x^n)$ in the process of making n arbitrarily large. It is simple to note that $k(\varepsilon, \mu_x^n)$ grows with n and it is upper bounded by $|\mathcal{A}|^n$, therefore $\log_2 k(\varepsilon, \mu_x^n) \leq n \log_2 |\mathcal{A}|$ and, consequently, taking the limits on the number of simulations the basic upper bound is

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 k(\varepsilon, \mu_x^n) \leq \log_2 |\mathcal{A}|. \quad (\text{A.2})$$

This means that $k(\varepsilon, \mu_x^n)$ grows (with n) at most exponentially with an exponential rate given by $\log_2 |\mathcal{A}|$. The next result in (A.3) stipulates that the precise exponential rate of $k(\varepsilon, \mu_x^n)$ (as n goes to infinity) approaches $\mathcal{H}(\mu_x)$, when making ε arbitrary small, in the sense that

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 k(\varepsilon, \mu_x^n) = \mathcal{H}(\mu_X). \quad (\text{A.3})$$

A.1.2 Derivation of The Representation Result in (A.3)

To show the identity in (A.3), it is useful to introduce a stronger notion of *achievable rate* for the entire i.i.d. process $\{\mu_X^n : n \geq 1\}$ that is commonly used in the context of almost lossless fixed-rate source coding (Cover and Thomas, 2006, Chap.3) and (Yeung, 2002, Chap.4).

Definition 4 For the i.i.d. simulation of μ_X , the rate $r > 0$ is said to be achievable for $\{\mu_X^n : n \geq 1\}$, if there is a sequence of sets $\{B_n \subset \mathcal{A}^n : n \geq 1\}$ that captures all the probability in the stronger sense that

$$\lim_{n \rightarrow \infty} \mu_X^n(B_n) = 1,$$

and

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 |B_n| \leq r.$$

Definition 5 In this context, the minimum achievable rate for $\{\mu_X^n : n \geq 1\}$ is

$$R^*(\mu_X) \equiv \min \{r : r > 0 \text{ is achievable for } \{\mu_X^n : n \geq 1\}\}.$$

From the Definitions, it is direct to show that $\limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 k(\varepsilon, \mu_x^n) \leq R^*(\mu_X)$ for all $\varepsilon > 0$ and consequently,

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 k(\varepsilon, \mu_x^n) \leq R^*(\mu_X). \quad (\text{A.4})$$

Remarkably for the i.i.d. context, it is well-known that the limit in the left-hand side of (A.4) is well-defined and matches $R^*(\mu_X)$. Furthermore, this expression reduces to the entropy of μ_x . This result was proved in the original paper of Shannon (Shannon, 1948) and it is highlighted in the following result:

THEOREM 1 (Shannon, 1948) $\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 k(\varepsilon, \mu_x^n) = R^*(\mu) = \mathcal{H}(\mu_X)$.

Therefore, the main conclusion of this analysis is that the Shannon entropy determines the complexity of simulating a finite alphabet probability in the precise operational sense defined in Definitions 3 and 5.

The proof of Theorem 1 is a direct consequence of the celebrated (weak) *asymptotic equipartition property* (AEP), first stated by Shannon and proved for the i.i.d. case in (Shannon, 1948). A systematic and clear exposition of this property and the proof of Theorem 1 can be found in (Cover and Thomas, 2006, Chap.3) and (Yeung, 2002, Chap.4) in the context of what is known in information theory as the *Shannon source coding theorem*. To conclude this part, it is important to elaborate some observations and implications about Theorem 1:

1. This result shows that there is a collection of sequences $\{B_n \subset \mathcal{A}^n : n \geq 1\}$ that captures asymptotically all the probability, that is, the collection is typical in the sense that $\lim_{n \rightarrow \infty} \mu_X^n(B_n) = 1$, and its cardinality grows exponentially at a rate that is precisely the entropy of μ_X , that is, as n goes to infinity $|B_n| \approx 2^{n \cdot \mathcal{H}(\mu_x)}$.
2. If $\mathcal{H}(\mu_X) < \log_2 |\mathcal{A}|$, then the collection of typical sequences, mentioned in the previous point, is an arbitrary small fraction of all the possible sequences, in the sense that

$$\frac{|B_n|}{|\mathcal{A}^n|} \approx \frac{2^{n \cdot \mathcal{H}(\mu_x)}}{2^{n \cdot \log_2 |\mathcal{A}|}} = 2^{-n \cdot (\log_2 |\mathcal{A}| - \mathcal{H}(\mu_x))} \longrightarrow 0. \quad (\text{A.5})$$

Note that the cardinality ratio in (A.5) goes to zero exponentially with n at a rate given by $\log_2 |\mathcal{A}| - \mathcal{H}(\mu_x) > 0$. Then for a i.i.d process $\{\mu_X^n : n \geq 1\}$ with entropy strictly smaller than $\log_2 |\mathcal{A}|$ there is a tiny fraction of sequences that characterize the i.i.d. process induced by μ_X . Here, $\log_2 |\mathcal{A}|$ is the maximum entropy only achieved by the uniform distribution on \mathcal{A} (Cover and Thomas, 2006).

3. For the achievability part of this result, Shannon proposes a specific collection of

typical sequences given by

$$B_n(\varepsilon) = \left\{ (x_1, \dots, x_n) \in \mathcal{A}^n : \left| -\frac{1}{n} \log_2 \mu_X^n(x_1, \dots, x_n) - \mathcal{H}(\mu_X) \right| \leq \varepsilon \right\}. \quad (\text{A.6})$$

that as n goes to infinity has the following properties (Cover and Thomas, 2006):

(a) it is a typical set: $\lim_{n \rightarrow \infty} \mu_X^n(B_n(\varepsilon)) = 1$.

(b) if $(x_1, \dots, x_n) \in B_n(\varepsilon)$ then

$$2^{-n \cdot (\mathcal{H}(\mu_X) + \varepsilon)} \leq \mu_X^n(x_1, \dots, x_n) \leq 2^{-n \cdot (\mathcal{H}(\mu_X) - \varepsilon)}.$$

(c) $(1 - \varepsilon)2^{n \cdot (\mathcal{H}(\mu_X) - \varepsilon)} \leq |B_n(\varepsilon)| \leq 2^{n \cdot (\mathcal{H}(\mu_X) + \varepsilon)}$.

Here, Achievability refers to the construction of a sequence that is typical and achieves a rate smaller or equal to the entropy (Cover and Thomas, 2006). Thus, the relevant aspect of this construction is the fact that as n progresses to infinity the elements of this typical set become uniformly distributed. Considering ε sufficiently small, for all $(x_1, \dots, x_n) \in B_n(\varepsilon)$, then:

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

Then within this set $B_n(\varepsilon)$, 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. Consequently from this fact, it is clear that the size of this typical set plays a major role from the point of view of specifying the complexity of the simulation task, which is formalized on the concept of minimum achievable rate in Definition 5, and that is precisely given by $\mathcal{H}(\mu_X)$.

A.2 App. II: Proof of Proposition 1

PROOF. After obtaining the sequential rule $\tilde{f}_k^* \in \mathbf{F}_k$ to solve the adaptive sensing rule of size k , the remaining posterior uncertainty is given by

$$H(\hat{X}_{\tilde{f}_k^*} | X_{\tilde{f}_k^*}). \quad (\text{A.7})$$

and, the reduction on the uncertainty, concerning to the information of \tilde{f}_k^* to resolve \bar{X} is given by

$$I(\tilde{f}_k^*) = H(\bar{X}) - H(\hat{X}_{\tilde{f}_k^*} | X_{\tilde{f}_k^*}) = H(X_{\tilde{f}_k^*}) \geq 0. \quad (\text{A.8})$$

As the sets $\hat{X}_{\tilde{f}_k^*}$ and $X_{\tilde{f}_k^*}$ conform a partition of \bar{X} ,

$$H(\bar{X}) = H(\hat{X}_{\tilde{f}_k^*}, X_{\tilde{f}_k^*}). \quad (\text{A.9})$$

In addition, by the chain rule of entropy

$$H(\bar{X}) = H(X_{\tilde{f}_k^*}) + H(\hat{X}_{\tilde{f}_k^*}, X_{\tilde{f}_k^*}). \quad (\text{A.10})$$

Applying (A.9) and (A.10), the information gain of the rule \tilde{f}_k^* can be rewritten by

$$\begin{aligned} I(\tilde{f}_k^*) &= H(X_{\tilde{f}_k^*}) + H(\hat{X}_{\tilde{f}_k^*}, X_{\tilde{f}_k^*}) - H(\hat{X}_{\tilde{f}_k^*} | X_{\tilde{f}_k^*}) \\ &= H(X_{\tilde{f}_k^*}) \\ &= H(X_{\tilde{f}_k^*(1)}, X_{\tilde{f}_k^*(2)}, \dots, X_{\tilde{f}_k^*(k)}) \\ &= H(X_{(i_1^*, j_1^*)}, X_{(i_2^*, j_2^*)}, \dots, X_{(i_k^*, j_k^*)}) \end{aligned} \quad (\text{A.11})$$

The expression (A.11) can be rearranged by using the chain rule as:

$$I(\tilde{f}_k^*) = H(X_{(i_1^*, j_1^*)}) + \sum_{h=2}^k H(X_{(i_h^*, j_h^*)} | X_{(i_1^*, j_1^*)}, \dots, X_{(i_{h-1}^*, j_{h-1}^*)}). \quad (\text{A.12})$$

Finally, using (A.12), the information gain difference between the sampling steps k and $k-1$ is given by

$$\begin{aligned} I(\tilde{f}_k^*) - I(\tilde{f}_{k-1}^*) &= H(X_{(i_1^*, j_1^*)}) + \sum_{h=2}^k H(X_{(i_h^*, j_h^*)} | X_{(i_1^*, j_1^*)}, \dots, X_{(i_{h-1}^*, j_{h-1}^*)}) \\ &\quad - H(X_{(i_1^*, j_1^*)}) - \sum_{h=2}^{k-1} H(X_{(i_h^*, j_h^*)} | X_{(i_1^*, j_1^*)}, \dots, X_{(i_{h-1}^*, j_{h-1}^*)}) \\ &= H(X_{(i_k^*, j_k^*)} | X_{(i_1^*, j_1^*)}, \dots, X_{(i_{k-1}^*, j_{k-1}^*)}). \end{aligned} \quad (\text{A.13})$$

□

A.3 App. III: Optimality of the Sequential Rule for Fields with no Inter-Pixel Dependency

PROOF. For a field \bar{X} with no inter-pixel dependency, every random variable $X_{i,j}$ is statistically independent of others variables in the sense that $H(X_{i_1, j_1} | X_{i_2, j_2}) = H(X_{i_1, j_1})$, for all $(i_2, j_2) \in [M] \times [M] \setminus \{(i_1, j_1)\}$. Then, by the independence bound on entropy (theorem 2.6.6 in (Cover and Thomas, 2006)), the posterior field entropy associated with the application of the optimal rule f_k^* posted in (3.18) is given by,

$$H(X_f) = \sum_{(i,j) \in f_k^*} H(X_{i,j}) \quad (\text{A.14})$$

Thus, for a field with no inter-pixel dependency (nipd), the optimal sampling rule of size k can be state as

$$f_k^{*,nipd} = \arg \max_{f \in \mathbf{F}_k} \sum_{(i,j) \in f} H(X_{i,j}), \quad (\text{A.15})$$

which, by the non negativity of the entropy, is the problem of choosing the k positions with the highest a priori entropy.

In the case of the iterative sequential rule, the k -measurement is now given by,

$$(\mathbf{i}_k^{*,nipd}, \mathbf{j}_k^{*,nipd}) = \arg \max_{(i,j) \in [M] \times [M] \setminus \{(i_l^{*,nipd}, j_l^{*,nipd}) : l=1, \dots, k-1\}} H(X_{i,j}), \quad (\text{A.16})$$

which correspond to choose the location with the k -th highest a priori entropy. Therefore, the sequential rule $\tilde{f}_k^{*,nipd} \in \mathbf{F}_k$ can be obtained by

$$\begin{aligned} \tilde{f}_k^{*,nipd}(1) &= (\mathbf{i}_1^{*,nipd}, \mathbf{j}_1^{*,nipd}), \\ \tilde{f}_k^{*,nipd}(2) &= (\mathbf{i}_2^{*,nipd}, \mathbf{j}_2^{*,nipd}), \\ &\dots, \\ \tilde{f}_k^{*,nipd}(k) &= (\mathbf{i}_k^{*,nipd}, \mathbf{j}_k^{*,nipd}). \end{aligned} \quad (\text{A.17})$$

Finally, by construction, the optimal combinatorial rule is equal to the sequential approach under the assumption of statistical independence,

$$\tilde{f}_k^{*,nipd} = f_k^{*,nipd}. \quad (\text{A.18})$$

meaning that in this scenario the optimal sampling can be achieved by the sequential approach. This simple case summarizes the nature of the optimal sampling approach that tries to resolve the locations with highest uncertainty in order to increase the average knowledge of the global field. As the inter-pixel dependency increases and the multiple-point statistics of the field becomes more complex, to take into account the high order conditional entropies becomes essential to avoid locations with redundant information. □

A.4 App. IV: Analysis of the Adaptive Sampling Scheme for the Markov Chain Case

Here, the *AMIS* $\{ \tilde{f}_k^a(\cdot|\cdot) : k \in [N] \}$ in (3.36) is compared with the *SMIS* approach $\{ \tilde{f}_k^*(\cdot) : k \in [N] \}$ in (3.28). This comparison is performed in the context of a finite length one-dimensional Markov source presented in Sect. 3.6 in terms of field resolvability (resolution of uncertainty) and the estimation of non-sensed positions. To evaluate the quality on resolving non-sensed position from the information of sensed position, the conditional entropy conditioning on the data is used, given a sensed rule f_k and its sensed data $x_f = (x_1, \dots, x_{k-1}) \in \mathcal{A}^{k-1}$,

$$H(\hat{X}_{f_k} | X_f = x_f). \quad (\text{A.19})$$

By extending the result in Proposition 4, there is a simple algorithm to compute (A.19) by the Markov assumption not reported here for the space constraint.

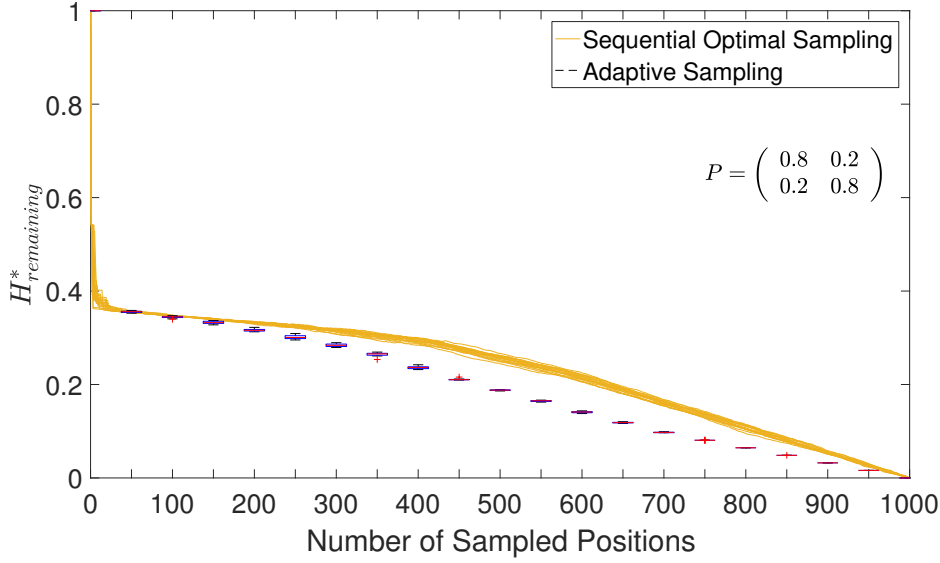


Figure A.1: Remaining conditional entropy by considering the previous sampled locations and its measurements. Symmetric transition matrix ($\beta = 0.2$)

Also the problem of estimating non-sensed position from the sensed data is considered for which the minimum mean square error estimator (MMSE) is applied given by the conditional mean (Gray and Davisson, 2004). Given f_k and its sensed data $x_f = (x_1, \dots, x_{k-1}) \in \mathcal{A}^{k-1}$, the MMSE estimator of \hat{X}_{f_k} from $X_f = x_f$ is given in closed form by $\mathbb{E}(\hat{X}_f | X_f = x_f) \in \mathcal{A}^{N-k}$, which is a function of x_f . This problem reduces to compute point-wise the expectation for every non-sensed position given $X_f = x_f$, which is a simple task under the Markov assumption.

A.5 App. V: Analysis of the Partial Update of Conditional Entropies

It can be argued that the use of the partial update of the conditional probabilities proposed in Sect. 3.7 can be justified in the context of SMIS in (3.27). Considering the stage $k-1$ and k of this algorithm, for an arbitrary unmeasured location, the focus is to evaluate the conditional entropy $H(X_{i,j} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-2}^*, j_{k-2}^*})$ for the stage $k-1$, and $H(X_{i,j} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-2}^*, j_{k-2}^*}, X_{i_{k-1}^*, j_{k-1}^*})$ for the stage k . Thus, the point-wise reduction of conditional entropy for the location (i, j) from the stage $k-1$ to the state k is given by

$$\begin{aligned}
 \Delta H^{(k-1) \rightarrow k}(X_{i,j}) &= H(X_{i,j} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-2}^*, j_{k-2}^*}) \\
 &\quad - H(X_{i,j} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-2}^*, j_{k-2}^*}, X_{i_{k-1}^*, j_{k-1}^*}) \\
 &= I(X_{i,j} | X_{i_1^*, j_1^*}, \dots, X_{i_{k-2}^*, j_{k-2}^*}; X_{i_{k-1}^*, j_{k-1}^*}) \tag{A.20}
 \end{aligned}$$

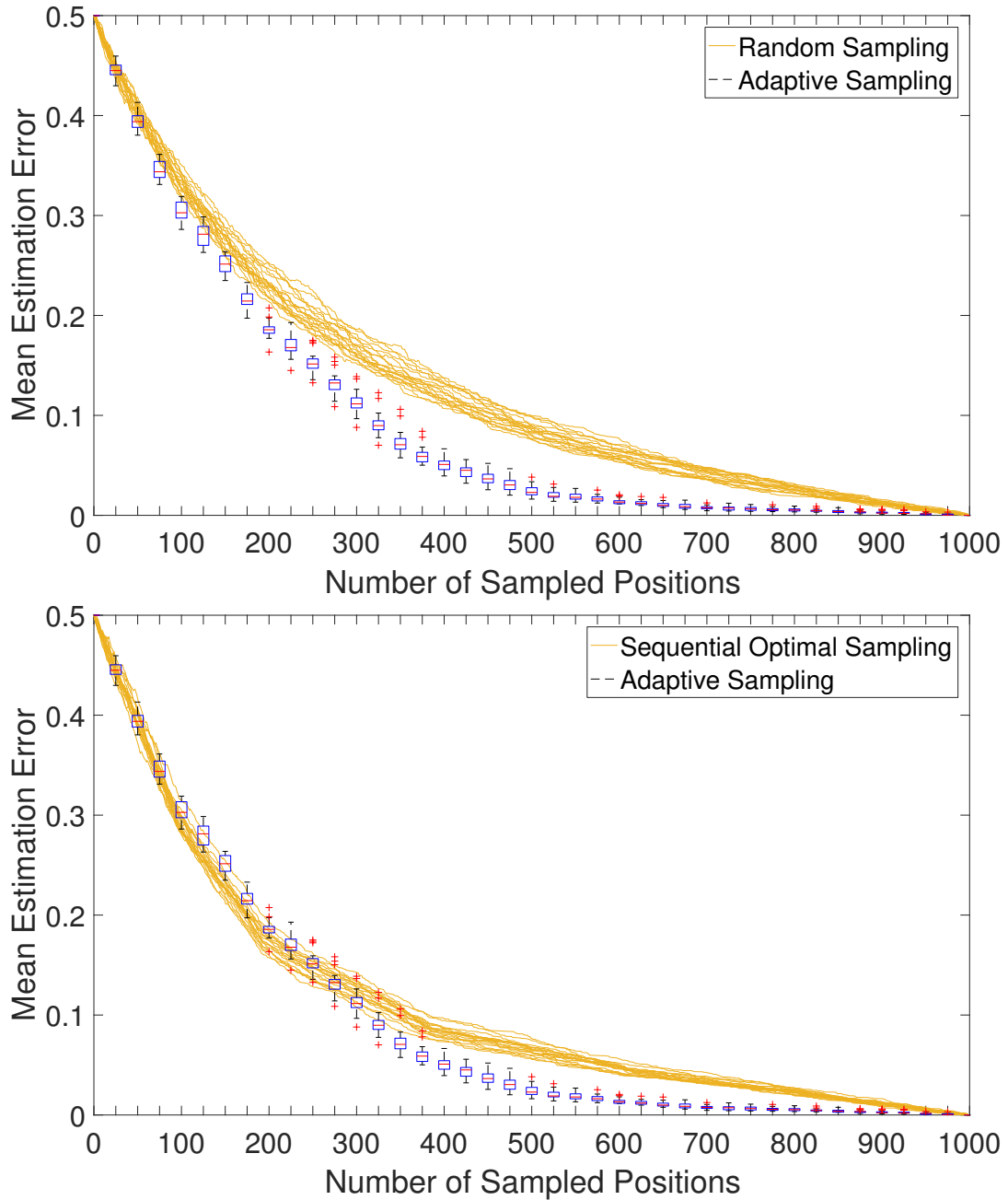


Figure A.2: Estimation error considering the previous sampled locations and its measurements. Symmetric transition matrix with $\beta = 0.1$. Top: Random Sampling vs AMIS Method, Bottom: SMIS vs AMIS

where, for $k = 2$ it is clear that the reduction in entropy corresponds exactly to mutual information

$$\begin{aligned}\Delta H^{(1)\rightarrow 2}(X_{i,j}) &= H(X_{i,j}) - H(X_{i,j}|X_{i_1^*,j_1^*}) \\ &= I(X_{i,j}; X_{i_1^*,j_1^*}).\end{aligned}\tag{A.21}$$

As the algorithm works in the search for new measurements, the mutual information is conditioned on previous measurements and, consequently, the approach used in the entropy map update deteriorates. However, recalculating the entropy maps every 20 samples has provided satisfactory results.

Appendix B

Complementary Material for Chapter 4

B.1 App. I: Pseudocode

The implemented pseudo-code that summarizes the proposed framework is shown here.

Algorithm 2: Pseudo code implementing the adaptive framework from rules in (4.8).

```

1 Initialization
2 Reference image  $\bar{X}$ , the regularization term  $\alpha$ , the batch size  $r$ , the threshold  $tsh_{grade}$  for
  ore-waste selection and the number of samples to take  $K$ 
  /* Output: set of sampled positions */
3  $f_K^{bench} = \emptyset$ 
  /* Inputs and variables.  $\mathcal{H}^k$  Estimated Remaining entropy for  $k - 1$  samples
  */
4  $\mathcal{H}^k = \text{ones}(\text{size}(\bar{X}))$ ,  $\mathcal{D} = \text{inf}(\text{size}(\bar{X}))$ ,  $\text{distMean} = \text{inf}$ 
5 Computation
6 for  $bench \leftarrow 1$  do
7    $f_K^1 = \text{Regular\_Sampling\_Approach}(K)$ 
8    $\hat{X}^1 = \text{Kriging\_From\_Samples}(f_K^1)$ 
9    $\hat{I}^1 = \hat{X}^1 \geq tsh_{grade}$ 
10 for  $bench \leftarrow 2$  to  $B$  do
11   /* Estimate initial TI from the upper bench */
12    $\hat{TI}^{bench} \leftarrow \hat{I}^{bench-1}$ 
13   /* Init the adaptive sampling process */
14    $f_K^{bench} = \emptyset$ ,  $f_K = \emptyset$ 
15   for  $k \leftarrow 1$  to  $K$  do
16      $f_{k-1}^{bench} \leftarrow f_K$ 
17     /* Update MPS simulations from previous available samples */
18     if  $\text{Criterion\_To\_Update\_MPS\_Realizations}(k, r)$  then
19        $\mathcal{H}^k =$ 
20          $\text{EntropyEstimated\_by\_MPS\_}\&\_Samples(\hat{TI}^{bench}, X_{f_{k-1}^{bench}}, \text{size}(\bar{X}^{bench}), f_{k-1}^{bench})$ 
21          $\mathcal{D} = \text{Estimated\_Distances\_From\_Sampled\_Locations}(\text{size}(\bar{X}), f_{k-1}^{bench})$ 
22          $\text{distMean} = \text{Mean}(\mathcal{D})$ 
23          $\hat{X}^{bench} = \text{Kriging\_From\_Samples}(f_{k-1}^{bench})$ 
24          $\hat{I}^{bench} = \hat{X}^1 \geq tsh_{grade}$ 
25          $\hat{TI}^{bench} \leftarrow \hat{I}^1$ 
26     else
27        $\mathcal{H}^{bench} = \mathcal{H}^{bench} - \text{Local\_Mutual\_Information}_{\text{empirical}}(f_{k-1}^{bench}(k - 1))$ 
28        $\mathcal{D} = \mathcal{D} \cdot$ 
29          $\text{Radial\_Attenuation\_Centred\_At\_Location}(\text{size}(\bar{X}^{bench}), \text{distMean}, k - 1)$ 
30     /* Set of current available position to sample. The complement set
31     of the previously sampled locations */
32      $\hat{f}_{k-1}^{bench} = \text{Complement\_Set\_Of\_F}(f_{k-1}^{bench})$ 
33     /* Regularized Criterion by Mixing of Entropy and Distance */
34      $\mathcal{M} = \alpha \cdot \mathcal{H}^{bench} + (1.0 - \alpha) \cdot \mathcal{D}$ 
35     /* Choose a location randomly from the set of non sampled locations
36     with maximal value for the objective function */
37      $f_K^{bench}(k) = \text{Select\_Random\_Location\_From\_Maximum\_Criterion}(\mathcal{M}_{\hat{f}_{k-1}^{bench}})$ 

```

B.2 App. II: Detailed Database

B.2.1 Case Study 2

For case study 2, the available drill-hole samples and blasting samples for the 6 benches are shown in fig. B.1 and fig. B.2 respectively. The ground truth was created using the software kt3d (fig. B.3).

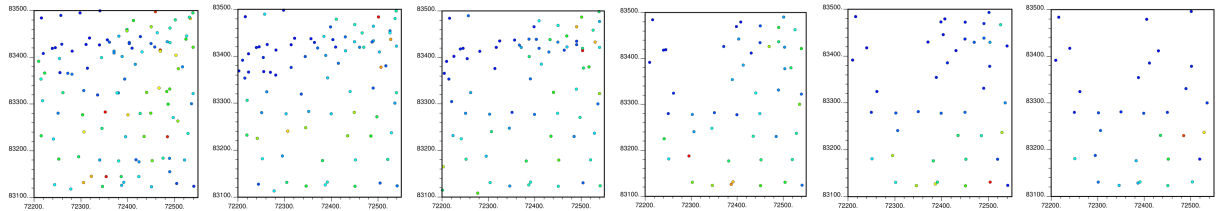


Figure B.1: Drill-hole samples data for case study 2. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

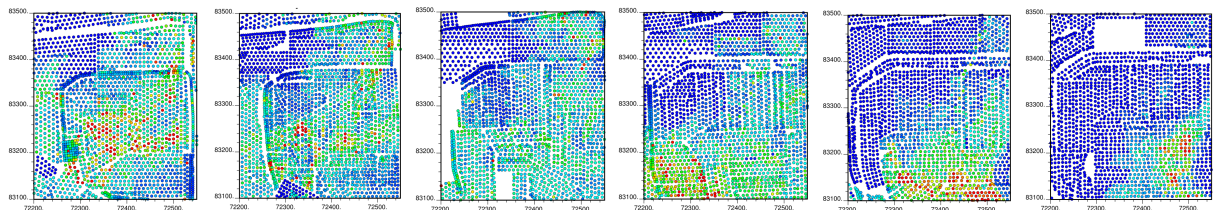


Figure B.2: Blast-hole data for case study 2. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

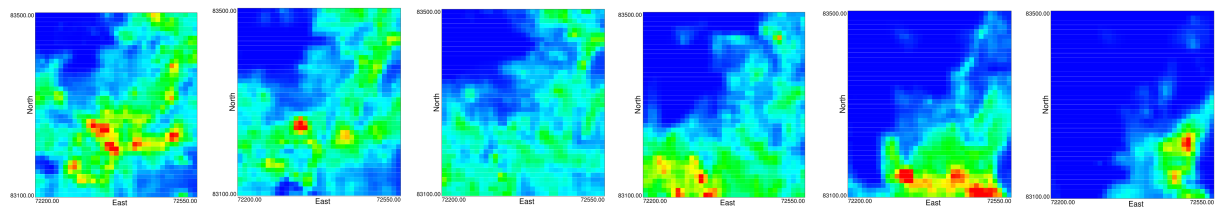


Figure B.3: Ground truth estimated from drill-holes and blast-holes for case study 2. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

B.2.2 Case Study 3

The 6 benches for case study 3 are shown in the fig. B.4 and the fig. B.5 describing the available drill-holes and blast-holes, respectively. The reference dense benches created by kriging analysis are shown in fig. B.6.

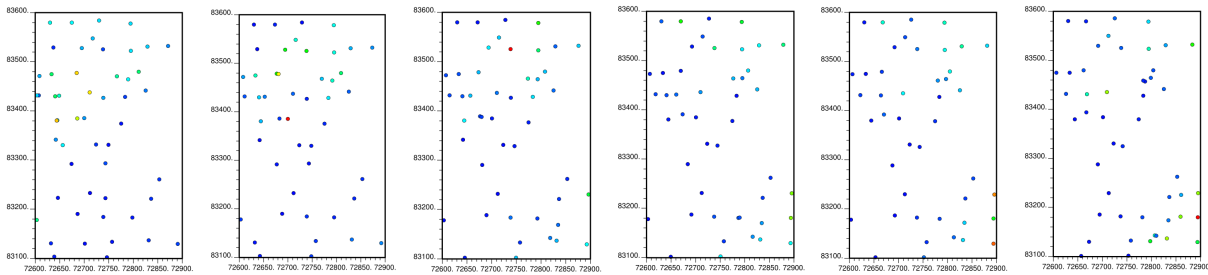


Figure B.4: Drill-hole samples data for case study 3. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

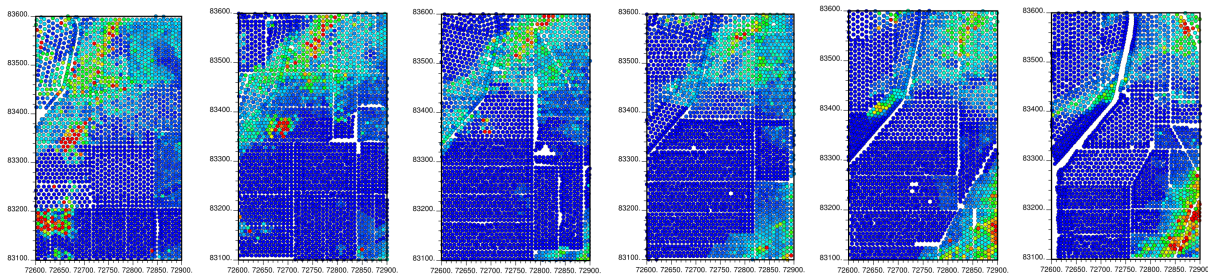


Figure B.5: Blast-hole data for case study 3. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

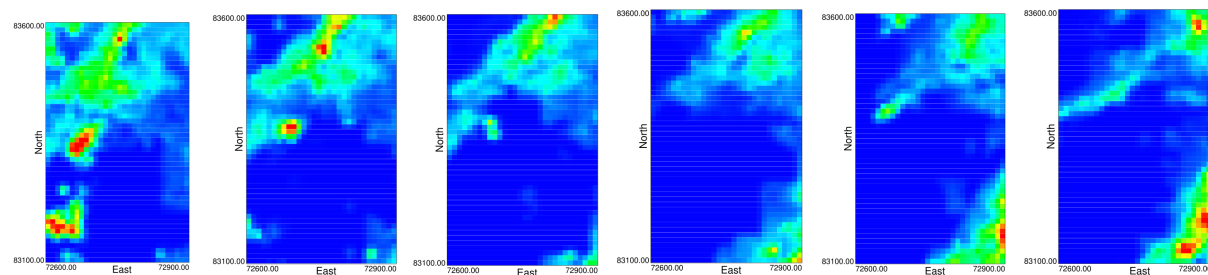


Figure B.6: Ground truth estimated from drill-holes and blast-holes for case study 3. From left to right: Benches 1-6. Colormap: the same as in Fig. 4.3.

B.3 App. III: Additional Experimental Results

Case study 1: *Cut-Off* grade 1.241%

With *Cut-Off* grade 1.241%, Figs. B.7, B.8 and B.9 describe the outcome.

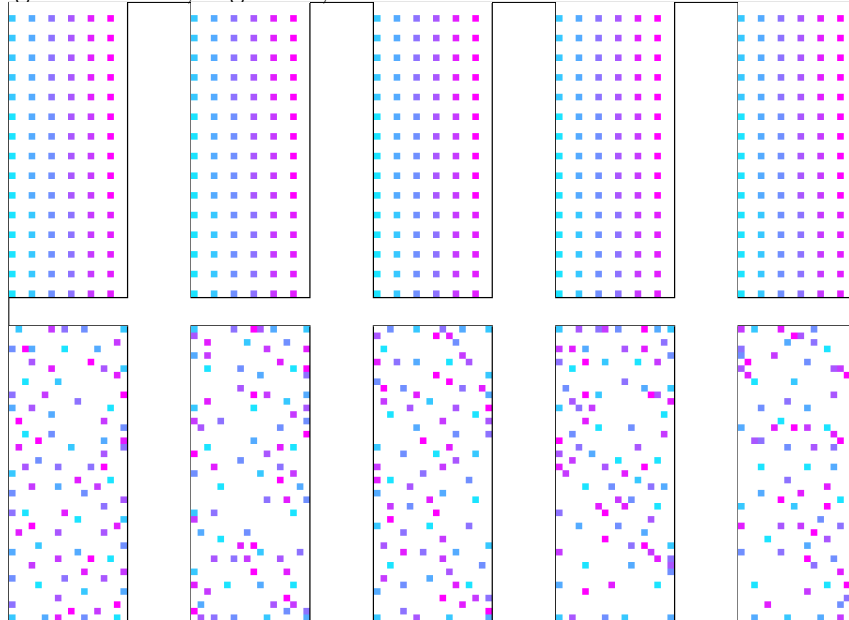


Figure B.7: Samples for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade 1.241%.

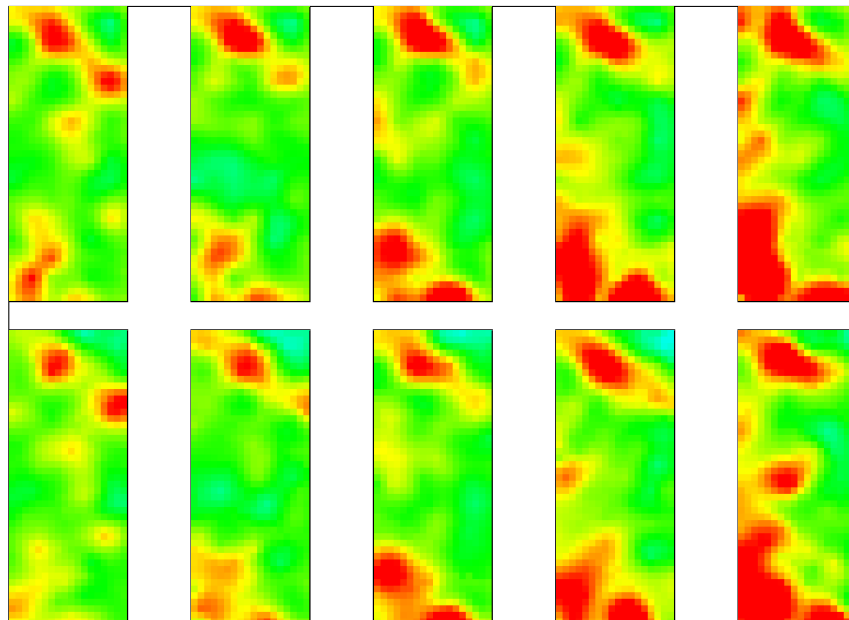


Figure B.8: Estimated grade for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade 1.241%.

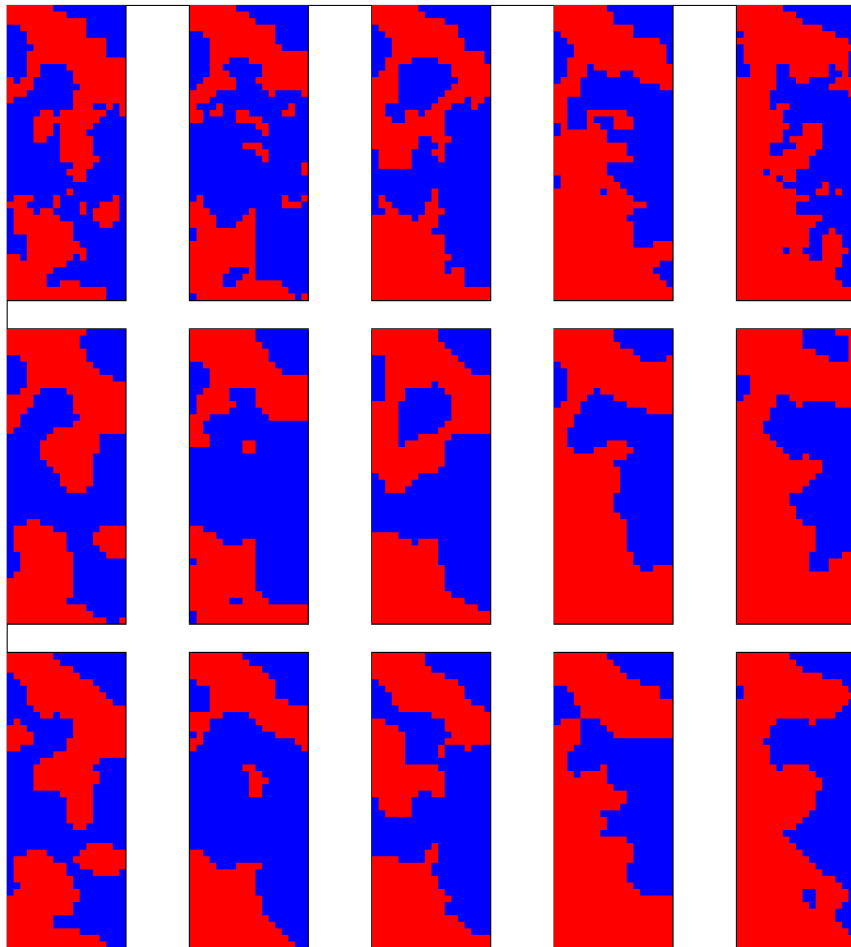


Figure B.9: Estimated grade control for Case Study 1. From left to right: Benches 2-6. From top to bottom: Ground truth, structured sampling, adaptive sampling using *Cut-Off* grade 1.241%.

		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
Actual class	Ore	89.5%	10.5%	82.8%	17.2%	89.9%	10.1%	93.2%	6.8%	93.1%	6.9%
	Waste	15.9%	84.1%	9.8%	90.2%	9.2%	90.8%	14.0%	86.0%	23.4%	76.6%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	
		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
Actual class	Ore	91.9%	8.1%	89.4%	10.6%	92.9%	7.1%	95.2%	4.8%	92.4%	7.6%
	Waste	14.5%	85.5%	10.2%	89.8%	11.1%	88.9%	14.9%	85.1%	16.8%	83.2%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

Figure B.10: Confusion Matrix for Case Study 1. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 1.241%.

Case study 1: *Cut-Off* grade 1.518%

Considering the *Cut-Off* grade 1.518%, the outcome for the benches 2,3,4,5 and 6 is described in Figs. B.11, B.12 and B.13.

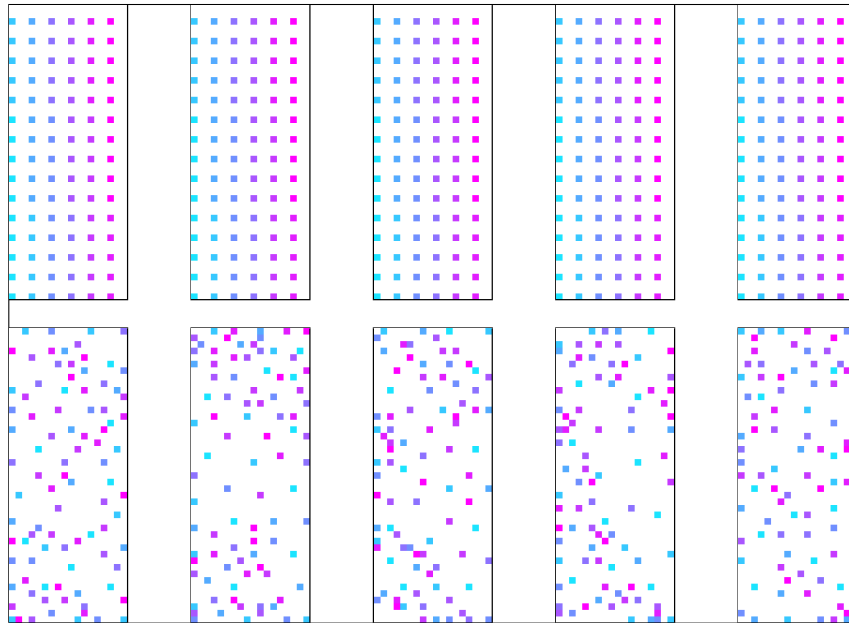


Figure B.11: Samples for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade 1.518%.

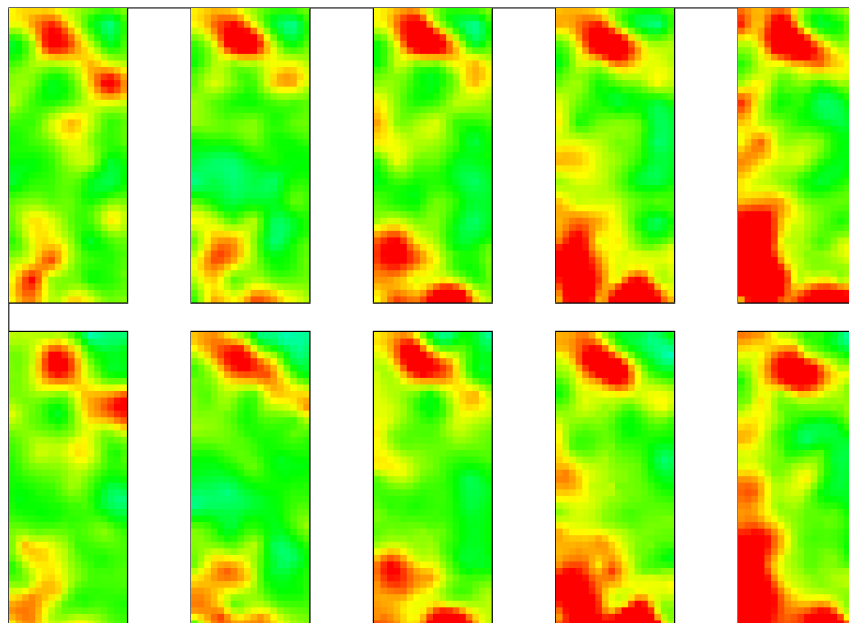


Figure B.12: Estimated grade for Case Study 1. From left to right: Benches 2-6. Top: Kriging from structured sampling. Down: Kriging from adaptive sampling using *Cut-Off* grade 1.518%.

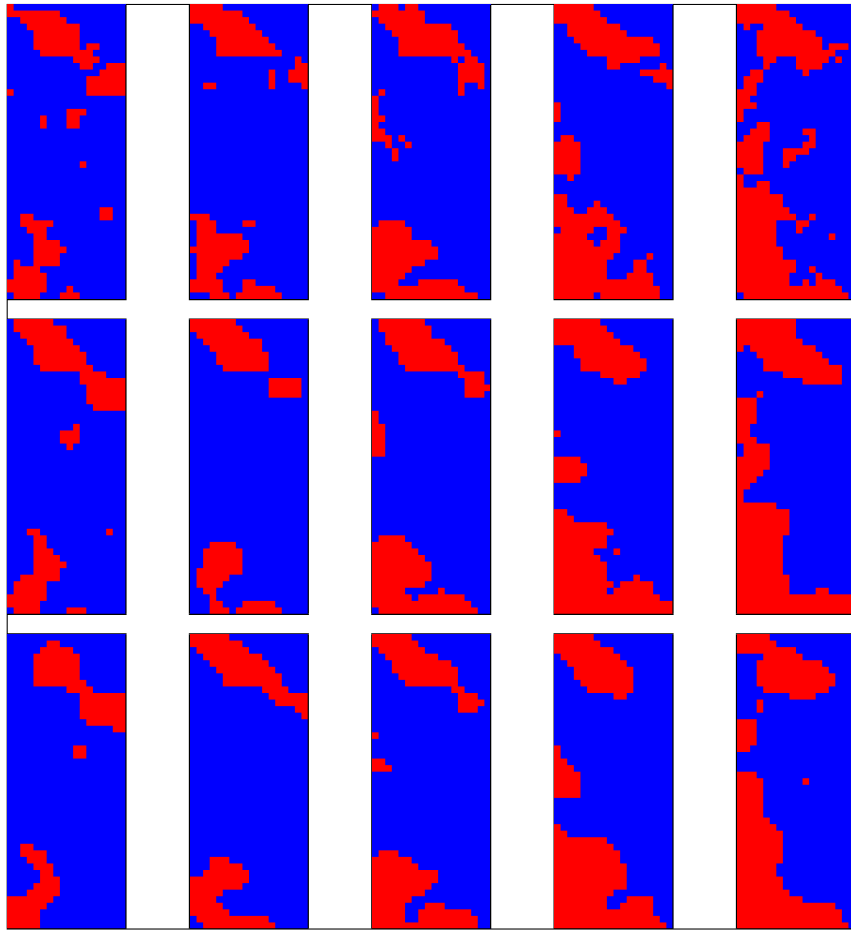


Figure B.13: Estimated grade control for Case Study 1. From left to right: Benches 2-6. From top to bottom: Ground truth, structured sampling, adaptive sampling using *Cut-Off* grade 1.518%.

		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
Actual class	Ore	75.0%	25.0%	75.8%	24.2%	85.0%	15.0%	84.3%	15.7%	86.0%	14.0%
	Waste	3.2%	96.8%	2.2%	97.8%	2.7%	97.3%	5.7%	94.3%	9.0%	91.0%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	
		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
Actual class	Ore	84.4%	15.6%	81.4%	18.6%	88.4%	11.6%	86.6%	13.4%	84.0%	16.0%
	Waste	3.5%	96.5%	4.3%	95.7%	3.2%	96.8%	5.9%	94.1%	6.1%	93.9%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

Figure B.14: Confusion Matrix for Case Study 1. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 1.518%.

Confusion Matrices for Case Study 2

		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
True class	Ore	98.9%	1.1%	98.0%	2.0%	97.9%	2.1%	95.9%	4.1%	95.1%	4.9%
	Waste	17.6%	82.4%	9.6%	90.4%	6.9%	93.1%	6.8%	93.2%	2.6%	97.4%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
True class	Ore	99.2%	0.8%	98.9%	1.1%	97.7%	2.3%	96.9%	3.1%	97.9%	2.1%
	Waste	16.6%	83.4%	9.6%	90.4%	5.6%	94.4%	4.5%	95.5%	1.4%	98.6%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

Figure B.15: Confusion Matrix for Case Study 2. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 0.22%.

		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
True class	Ore	93.3%	6.7%	94.1%	5.9%	90.7%	9.3%	96.3%	3.7%	90.5%	9.5%
	Waste	13.5%	86.5%	15.7%	84.3%	6.7%	93.3%	3.5%	96.5%	1.5%	98.5%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

		Bench 2		Bench 3		Bench 4		Bench 5		Bench 6	
True class	Ore	95.7%	4.3%	93.5%	6.5%	92.8%	7.2%	95.8%	4.2%	95.8%	4.2%
	Waste	15.3%	84.7%	12.7%	87.3%	6.2%	93.8%	1.7%	98.3%	0.6%	99.4%
		Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

Figure B.16: Confusion Matrix for Case Study 2. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 0.445%.

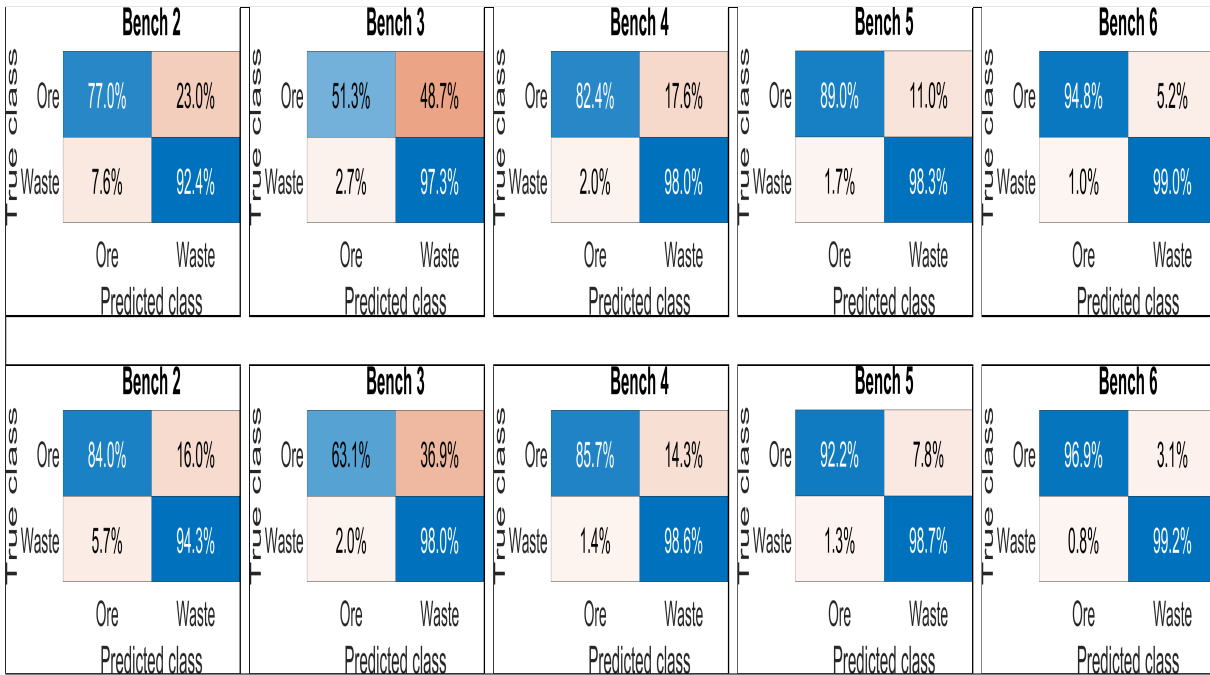


Figure B.17: Confusion Matrix for Case Study 2. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 0.692%.

Confusion Matrices for Case Study 3



Figure B.18: Confusion Matrix for Case Study 3. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 0.115%.

	Bench 2		Bench 3		Bench 4		Bench 5		Bench 6			
True class	Ore	93.0%	7.0%	Ore	88.4%	11.6%	Ore	93.0%	7.0%	Ore	89.7%	10.3%
	Waste	5.5%	94.5%	Waste	3.6%	96.4%	Waste	4.7%	95.3%	Waste	3.0%	97.0%
	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
	Predicted class		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	
	Bench 2		Bench 3		Bench 4		Bench 5		Bench 6			
True class	Ore	94.5%	5.5%	Ore	95.6%	4.4%	Ore	94.6%	5.4%	Ore	89.7%	10.3%
	Waste	3.2%	96.8%	Waste	2.5%	97.5%	Waste	2.6%	97.4%	Waste	1.6%	98.4%
	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
	Predicted class		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

Figure B.19: Confusion Matrix for Case Study 3. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 0.273%.

	Bench 2		Bench 3		Bench 4		Bench 5		Bench 6			
True class	Ore	81.9%	18.1%	Ore	70.1%	29.9%	Ore	75.0%	25.0%	Ore	84.9%	15.1%
	Waste	4.6%	95.4%	Waste	2.2%	97.8%	Waste	2.8%	97.2%	Waste	1.3%	98.7%
	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
	Predicted class		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	
	Bench 2		Bench 3		Bench 4		Bench 5		Bench 6			
True class	Ore	79.4%	20.6%	Ore	71.8%	28.2%	Ore	80.2%	19.8%	Ore	89.5%	10.5%
	Waste	2.6%	97.4%	Waste	1.8%	98.2%	Waste	1.3%	98.7%	Waste	2.2%	97.8%
	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste	Ore	Waste
	Predicted class		Predicted class		Predicted class		Predicted class		Predicted class		Predicted class	

Figure B.20: Confusion Matrix for Case Study 3. From left to right: Benches 2-6. Top: Structured sampling. Down: Adaptive sampling using *Cut-Off* grade 0.486%.

B.4 App. IV: Economic Evaluation

In order to perform a realistic evaluation of cost and profit several considerations must be defined for the mining and waste processing. In particular, in this work the variables summarized in table B.1 have been considered, where a range of realistic values for these variables is proposed. This analysis takes into account the block size (fixed to 1000 m^3 , the same block size used in the experimental section), the mining cost by mined ton, the metallurgical recovery, and the stripping ratio (the proportion of tons of expected waste material and ore material). For the purpose of the present analysis, the economic costs considered were the processing cost by processed ton, the price and selling cost per pound of copper.

Table B.1: Variables Considered in the Profit Analysis.

Variable	Symbol	Initial Value	Units	Range
Block size	$size_b$	1000	m^3	-
Density	d_b	2.70	t/m^3	2.3 - 2.9
Mining cost	$Cost^m$	3.0	$US\$/t_{mined}$	1.70 - 3.50
Stripping ratio	S_r	2.50	-	2.50 - 3.00
Processing cost	$Cost^p$	10.00	$US\$/t_{processed}$	7.00 - 12.00
Price of a lb of Cu	$Price_{Cu}$	2.30	$US\$/lb$	2.50 - 2.10
Selling cost by lb_{Cu}	$Cost^s$	0.50	$US\$/lb$	0.30 - 0.55
Metallurgical recovery	R_m	85.00	%	80.00 - 90.00

In practice, the *CutOff* grade, Cg , can be defined by,

$$Cg = \frac{10000 \cdot (Cost^m \cdot (1 + S_r) + Cost^p)}{(Price_{Cu} - Cost^s) \cdot R_m \cdot 2204.6}, \quad (B.1)$$

where the value 2204.6 corresponds to the conversion factor from pounds to tons.

Given the set of values for the considered parameters, then it is possible to estimate the profit of a block. In the one hand, for a block with an estimated grade under the *Cut-Off* grade value, for simplification it has been defined that the cost to process the waste block in the dump facilities is considered equal for each dumped block without taking into account its actual mineral grade or another variables. For this brief analysis, the revenue from dumping the block is zero (in practice if the block has a zero profit, since it has already been mined then it could be considered as stock pile). Thus:

$$Profit_d = -Cost^m \quad (B.2)$$

In the other hand, in the case of a block estimated as ore (estimated block grade is above the *Cut-Off* grade), then it is processed by the mine and its benefit is calculated considering the content of metal in percentage of tons of copper as,

$$Cont_m = \frac{Grade_b}{100.0} \cdot size_b \cdot d_b \quad (B.3)$$

The content of recovered metal in tons of copper is estimated as

$$Cont_m^r = \frac{Cont_m \cdot R_m}{100.0} \quad (B.4)$$

Then, the revenue from mining the block is estimated by,

$$Rev_b^m = Cont_m^r \cdot 2204.6 \cdot (Price_{Cu} - Cost^s), \quad (B.5)$$

while the processing cost of the block is estimated as

$$Cost_b^p = size_b \cdot d_b \cdot (Cost^m + Cost^p). \quad (B.6)$$

Finally, the profit of the processed block is defined by,

$$Profit_b^p = Rev_b^m - Cost_b^p. \quad (B.7)$$

Therefore, from Eq. (B.2), Eq. (B.7) and the cost of processing ore and waste blocks, it is possible to estimate the profit or loss of any block.

Considering the *Cut-Off* grade of the experimental analysis and Eq. (B.1), the appropriate economic parameters have been estimated. The initial values shown in Table B.1 are provided as an example to obtain the *CutOff* grade as the first quartile in Case Study 1. For each case study and for every empirical *CutOff* grade, the best set of parameters has been estimated in order to perform the economic analysis.

Then, from the experimental data the profit of every block has been evaluated in the proposed scenarios for the sampling strategies under analysis. Given an estimated bench block model from a specific sampling strategy, then the economic profit can be calculated as the sum of the profit for each block conforming this bench.

Appendix C

Complementary Information and Development

The following sections summarize several additional information required to contextualize the information presented in this thesis, but not included in the main text for the sake of fluency for the reader.

C.1 Details on *MPS*

C.1.1 *MPS* and the role of Training Images

MPS incorporates a prior model by the use of a training image. It can be defined as a $2-D$ or $3-D$ valid realization of a field with the same structures that the target field (e.g. structures like channels, reefs, bars, dikes, oriented facies) representing the full range of possible shapes and its scales. *MPS* was proposed for going beyond *two-point* statistics (Guardiano and Srivastava, 1993) using training images to describe the full range of structures present in the geological field. Therefore, training images allow to model complex geological features and their connectivity. The training image must be considered as a conceptual representation of the geological field instead of an actual realization of it, providing only relative spatial information of the distribution of variables of interest in the field (Scheidt and Caers, 2009).

The generation and selection of training images is an important challenge for the *MPS* methodology. A classical option is the simulation of unconditioned realizations using object-based approaches where an expert user defines the facies shapes and dimensions of interest. In order to select an appropriate training image from a set of available ones, a consistency check has been proposed to compare and validate available well data (Strebelle and Zhang, 2004).

Due to the limited size of training images only the inference of a very reduced portion of the real *multiple-point* statistics is accurate. Therefore, statistics for large scale structures are usually ignored because this can lead to undesired discontinuities.

C.1.2 Statistics from *MPS* Process

It is important to note that training images are a source of expected patterns in the target field. For example, let $X = \{X_i : i \in \{1, \dots, N\}\}$ be a categorical field to be simulated, with z_0, \dots, z_r different states defining the alphabet of an individual variable, X_i . The more classical *MPS* process is sequential by producing one pixel at time. In particular, a random path is defined to explore all positions to be simulated in the field (excluding sensed data positions, if available). Then the position simulated at a given time becomes conditioning data for the positions to be explored later in the path sequence. More precisely, let X_j denotes an unsampled variable, a context based rule is required to define the c closest and most relevant context for X_j that is denoted by $X_S = \{X_{S_1} = x_{S_1}, \dots, X_{S_c} = x_{S_c}\}$. These selected variables are chosen from the initial available data and from previously simulated variables in the path. Then, the probability that X_j has the state z conditioning on the context X_S is estimated by the *Bayes* rule:

$$p(X_j = z | X_S = x_S) = \frac{p(X_j = z, X_S = x_S)}{p(X_S = x_S)}. \quad (\text{C.1})$$

In Eq. (C.1) $p(X_j = z, X_S = x_S)$ and $p(X_j = z)$ are estimated from the aforementioned training image. In particular, $p(X_S = x_S) = \frac{\#(x_S)}{N}$ with N the number of pixels of the training image (in this case equal to the size of the target field to be simulated), and $\#(x_S)$ is the number of occurrences of the specific data pattern x_S in the training image. The probability $p(X_j = z, X_S = x_S)$ can be estimated as a pattern with one additional conditioning. Then, the conditional probability associated to the occurrence of state z at location $\{j\}$ can be estimated by:

$$p(X_j = z, X_S = x_S) = \frac{\#(x_{S,j})}{N} \quad (\text{C.2})$$

with $\#(x_{S,j})$ the number of occurrences of the pattern including the conditioning data and the central variable X_j with the specific value z . Thus, evaluating the expression (C.2) for the z possible values allows drawing a realization of the stochastic process for the explored variable.

C.2 Details of additional methods for Signal Recovery

C.2.1 Methods used in Sparse Signal Recovery

In this section we explore some available methods oriented to obtain a reconstruction of a signal from a small amount of measurements under assuming a sparse or compressible model for these signals.

Sparse Signal Recovery

The recovery formulations have a close relation with the inverse problem presented in Eq. (1.4), because the target is the reconstruction of the signal X from the observations $Y_{Sampled}$. In previous sections, signal reconstruction is not addressed because in the regime $m \ll N$ this is an ill-posed problem.

While the above is correct in general, recent results on *Compressed Sensing (CS)* theory provides novel insights for signal reconstruction under the assumptions of desired properties on the signal itself and in the sensing scheme, such as the sparsity at X and the incoherence at the sensing matrix \mathbf{A} (Candes et al, 2006a; Eldar and Kutyniok, 2012).

The *CS* theory has found applications on several areas such as signal representation, functional approximation, spectral estimation, cartography, medical imaging, speech signal processing, and sparse channel estimation (Elad, 2010; Eldar and Kutyniok, 2012; Starck et al, 2010).

Basic formulation

CS attempts to determine the minimal number of observations $Y_{Sampled}$ required to a stable reconstruction of a sparse¹ signal (with sparsity S). Considering a linear system, in Eq.(1.4) each individual observation is the inner product of the signal X , of size N , with a row vector in the sensing matrix \mathbf{A} as shown in Eq.(C.3).

$$Y_{Sampled} = \mathbf{A} \cdot X + \eta. \quad (\text{C.3})$$

The simplest approach for the noiseless case correspond to solve the ℓ_0 minimization problem:

$$(P_0) \quad \min_{X \in \mathbb{R}^N} \|X\|_0 \text{ subject to } Y_{Sampled} = \mathbf{A} \cdot X. \quad (\text{C.4})$$

For the noisy case:

$$(P_{0n}) \quad \min_{X \in \mathbb{R}^N} \|X\|_0 \text{ subject to } \|Y_{Sampled} - \mathbf{A} \cdot X\|_2 \leq \eta. \quad (\text{C.5})$$

The norm $\|\cdot\|_0$ counts the non-zero entries of the signal X while the norm $\|\cdot\|_2$ denotes *euclidean* norm.

The optimization problems formulated in Eq.(C.4) and Eq.(C.5) are *NP hard*. Due this fact, two main types of methods have been proposed in the last decade to achieve a practical solution.

In the one hand, greedy algorithms like *(Orthogonal) Matching Pursuit ((O)MP)* or *Thresholding* methods perform approximations to obtain a suboptimal solution. *Thresholding* method estimates the inner products of the target signal X with all sensing atoms² finding the largest

¹A signal is k -sparse in the canonical domain, if it has at most k terms different than zero

²Row vectors conforming the sensing system in the linear model in Eq.(C.3)

Table C.1: Summary of classical greedy approaches

Goal: An approximated estimation of X from $Y_{Sampled} = \mathbf{A} \cdot X$. A_j denotes columns of \mathbf{A} and $\mathbf{A}_\Lambda^\dagger$ the pseudo-inverse of \mathbf{A}_Λ

OMP	Thresholding
initialize: $R = Y_{Sampled}$, $\Lambda = \emptyset$ find: $k = \arg \max_j \langle r, A_j \rangle $ update: $\Lambda = \Lambda \cup \{i\}$, $R = Y_{Sampled} - \mathbf{A}_\Lambda \mathbf{A}_\Lambda^\dagger Y_{Sampled}$ iterate until R based stopping criterion output: $X = \mathbf{A}_\Lambda^\dagger Y_{Sampled}$	find: The collection Λ of indices providing largest sparse representation by $ \langle Y_{Sampled}, A_j \rangle $ output: $X = \mathbf{A}_\Lambda^\dagger \cdot Y_{Sampled}$

ones (in the absolute values) and finally calculating the orthogonal projection onto the span of the selected atoms. The *OMP* method is a sequential approach that selects the most representative atom (i.e. the one with largest absolute inner product with the signal residual) and estimates the signal by updating the residual (the misrepresented part of the observations) by the cumulative selection of atoms. See more details in Table C.1.

On the other hand, another alternative to address the (P_0) and (P_{0n}) is by a convex relaxation of the ℓ_0 norm. For the noiseless case, we obtain the relaxed problem termed Basis Pursuit (*BP*):

$$(P_1) \quad \min_{X \in \mathbb{R}^N} \|X\|_1 \text{ subject to } Y_{Sampled} = \mathbf{A} \cdot X. \quad (\text{C.6})$$

The noisy version of (P_1) is described by the expression termed *Basis Pursuit Denoising (BPDN)*:

$$(P_{1n}) \quad \min_{X \in \mathbb{R}^N} \|X\|_1 \text{ subject to } \|Y_{Sampled} - \mathbf{A} \cdot X\|_2 < \eta. \quad (\text{C.7})$$

Here, ℓ_1 -norm is calculated as $\|X\|_1 = \sum |X_i|$.

The main goal of the *CS* theory is to find guarantees of (near)perfect reconstruction and the associated sufficient conditions. Revolutionary results have established that under certain conditions greedy and *BP* approaches achieves guarantees for perfect reconstruction (Boyko et al, 2011; Candes et al, 2006a). For *BPDN*, it is required that the sensing matrix \mathbf{A} obeys a uniform uncertainty principle, which refers to the presence of well-conditioned submatrices in \mathbf{A} .

Formally, let $\Lambda \subset \{1, \dots, N\}$ be a collection of indices and the \mathbf{A}_Λ a submatrix of \mathbf{A} constructed using the columns of \mathbf{A} indexed by Λ . Then, the local isometry constant $\delta_\Lambda = \delta_\Lambda(\mathbf{A})$ is defined as the smallest value satisfying Eq. (C.8) for all vectors X supported on Λ .

$$(1 - \delta_\Lambda) \|X\|_2^2 \leq \|\mathbf{A}_\Lambda X\|_2^2 \leq (1 + \delta_\Lambda) \|X\|_2^2 \quad (\text{C.8})$$

Finally, the (global) restricted isometry property (*RIP*) is defined by the following constant:

$$\delta_S = \delta_S(\mathbf{A}) := \sup_{|\Lambda|=S} \delta_\Lambda(\mathbf{A}), \quad S \in \mathbb{N}. \quad (\text{C.9})$$

If \mathbf{A} has a small restricted isometry constant, i.e. $\delta_S(\mathbf{A}) \leq 1/2$, then \mathbf{A} satisfies a uniform uncertainty principle (Candes et al, 2006a).

An important theorem for *BPDN* states that if \mathbf{A} satisfies $\delta_{3S}(\mathbf{A}) + 3\delta_{4S}(\mathbf{A}) < 2$ for some $S \in \mathbb{N}$, and the noisy observations accomplish $Y_{\text{Sampled}} = \mathbf{A} \cdot X + \xi$ for some $\|\xi\|_2 \leq \eta$, then the signal $X^\#$ (solution of the problem (P_1n)) satisfies Eq. (C.10) for an appropriate value of C which only depends on δ_{3S} and δ_{4S} constants (Eldar and Kutyniok, 2012).

$$\|x^\# - x\|_2 \leq C\eta. \quad (\text{C.10})$$

In fact, if $\delta_{4S} \leq 1/3$ then $C \leq 15.41$. In addition, the formulation related with the Eq. (C.10) ensures exact reconstruction for the noiseless problem (P_1) with $\eta = 0$.

While these strong guaranties are only applicable for matrices satisfying the required principle, in practical cases few sensing schemes allow to reach perfect reconstruction. However, *CS* literature remarks that with high probability a $m \times N$ random matrix where the columns are drawn from distributions with certain concentration properties (such as *gaussian* distributions), would have small restricted isometry constants δ_S when $m = \mathcal{O}(S \log(N/S))$.

General Approach for Sparse Reconstruction

Conceptually, *CS* methodology poses a problem that arises quite naturally. The underlying idea corresponds to a regularized problem where the goal is to honor the available observations Y_{Sampled} while trying to find the simplest solution. In (P_0n) honoring of observations is carried out by bounding the *Euclidean* error while the simplicity of the solution is achieved by minimizing ℓ_0 -norm or ℓ_1 -norm.

It is possible to propose a more general scheme for this regularization approach. The compromise between fitting observations and simplicity can be subsumed by the following unconstrained convex minimization problem.

$$(P_G) \min_{X \in \mathbb{R}^N} \left\| \mathbf{C}_u^{\frac{-1}{2}} \cdot (Y_{\text{Sampled}} - \mathbf{A} \cdot X) \right\|_p + \gamma \cdot \|\mathbf{W} \cdot X\|_q. \quad (\text{C.11})$$

This regularization problem promotes simplicity using ℓ_q -norm on the target signal X . In addition, a weighting matrix W can be used to incorporate some prior information about the preponderance of some entries of X .

Honoring the observations can be achieved by the ℓ_p -norm of the observation error $Y_{\text{Sampled}} - \mathbf{A} \cdot X$. The use of the matrix $\mathbf{C}_u^{\frac{-1}{2}}$ allows working with signals that have some level of correlation. Approaches that consider the noisy scenario are usually termed as Noisy Compressive Sensing (*NCS*) approaches.

The problem (P_1n), known as Least-absolute shrinkage and selection operator (*LASSO*), is a version of (P_G) by ℓ_1 -norm regularization with a quadratic constraint (Boyko et al, 2011).

$$\min_{X \in \mathbb{R}^N} \gamma \|X\|_1 + \frac{1}{2} \cdot \|\mathbf{A} \cdot X - Y_{Sampled}\|_2^2. \quad (\text{C.12})$$

As previously presented, it approach finds the vector X with minimum ℓ_1 -norm that comes close to explaining the observations $Y_{Sampled}$.

An alternative approach called *Dantzig Selector* searches the minimum ℓ_1 -norm but with bounded residual correlation (Candes and Tao, 2007), as can be seen in the following expression:

$$\min_{X \in \mathbb{R}^N} \|X\|_1 \quad \text{subject to} \quad \left\| \mathbf{A}^\dagger \cdot (\mathbf{A} \cdot X - Y_{Sampled}) \right\|_\infty \leq \varepsilon. \quad (\text{C.13})$$

If the underlying target signal is a 2 - D image, an alternate recovery approach is promoting sparsity on gradients of the signal instead of the signal itself. The 2 - D gradient can be calculated from *total variations* (TV), where a possible definition is given by:

$$TV(X) := \sum \sqrt{(D_{h;ij} \cdot X)^2 + (D_{v;ij} \cdot X)^2} = \sum \|D_{ij} \cdot X\|_{2ij}. \quad (\text{C.14})$$

Where $D_{h;ij}$ denotes the gradients on horizontal orientation and $D_{v;ij}$ the vertical ones. Then, the total variations based *BPDN* (TV -*BPDN*) approaches is defined by:

$$\min_{X \in \mathbb{R}^N} TV(X) \quad \text{subject to} \quad \|Y_{Sampled} - \mathbf{A} \cdot X\|_2 < \eta. \quad (\text{C.15})$$

Finally, the *TV-Dantzig Selector* regularization problem can be written as:

$$\min_{X \in \mathbb{R}^N} TV(X) \quad \text{subject to} \quad \left\| \mathbf{A}^\dagger \cdot (\mathbf{A} \cdot X - Y_{Sampled}) \right\|_\infty \leq \varepsilon \quad (\text{C.16})$$

Although these combinations of regularization factors do not have known theoretical guarantees (as the case of *CS*), in practical applications of image processing they have demonstrated good performances (Chambolle and Caselles, 2010).

The rest of this chapter is dedicated to the detail of some results and formulations presented in this thesis

C.3 Resolvability Capacity of the random field X

It is interesting to analyze an indicator of the complexity of the field in terms of the capacity to resolve its uncertainty with K preferential measurements from rule f . Let define the resolvability capacity of X with K -measurements as:

$$\mathcal{C}_K \equiv \frac{I(f_K^*)}{H(X)} \in [0, 1], \quad (\text{C.17})$$

for all $K \in [N]$. This is the ratio between the information gain of the best sensor-placement rule and the whole entropy of the random field. The two extreme cases are $\mathcal{C}_K = 0$, which implies that the K -measurements produces no reduction in uncertainty, and $\mathcal{C}_K = 1$, which implies that there is no remaining uncertainty after taking the K -measurements, i.e., $H(X_{f_K^*}^* | X_{f_K^*}^*) = 0$. In general, \mathcal{C}_K has the following properties:

PROPOSITION 5 For any arbitrary random field X :

1. $\mathcal{C}_{K+1} \geq \mathcal{C}_K, \forall K \in \{1, \dots, N-1\}$.
2. $\mathcal{C}_N = 1$, and it can be defined that $\mathcal{C}_0 = 0$.

Hence, $\{\mathcal{C}_K : K \in [N]\}$ is a monotonically increasing sequence and its profile gives an insight of how simple (or complex) is to resolve the information of X in the process of taking K -optimal measurements.

C.3.1 Resolvability Capacity of the Iterated Principle

It is simple to show that the optimal solution f_K is better than the iterative solution \tilde{f}_K^* in the sense of information to resolve X . More precisely, for all $K \in \{1, \dots, N\}$

$$I(\tilde{f}_K^*) \leq I(f_K^*). \quad (\text{C.18})$$

Consequently, a concrete way to evaluate how much is lost in the reduction of the uncertainty between the combinatorial optimal scheme $\{f_K : K \geq 1\}$ and the iterative scheme $\{\tilde{f}_K^* : K \geq 1\}$, is given by the difference between \mathcal{C}_K in Eq. (C.17) and the resolvability capacity for the iterated solution \tilde{f}_K^* , given by:

$$\tilde{\mathcal{C}}_K \equiv \frac{I(\tilde{f}_K^*)}{H(X)} \in [0, 1]. \quad (\text{C.19})$$

In this thesis, it has been conjectured that the information loss $(\mathcal{C}_K - \tilde{\mathcal{C}}_K)_{K=1,\dots,N}$ is proportional to how much spatial dependency is presented in the joint distribution of field X . In one extreme, it is simple to prove that for a field with no inter-pixel (spatial) dependency, i.e., $P_X(X) = \prod_{i \in [N]} P_{X_i}(X_i)$, obtaining:

$$\mathcal{C}_K = \tilde{\mathcal{C}}_K, \quad (\text{C.20})$$

and furthermore, the iterative solution is optimal: $\tilde{f}_K^* = f_K^*$ for all K .

C.3.2 RAMIS and Resolvability Capacity Properties

In order to quantify the level of *knowledge* on the media given an optimal sampling $f_K^* = \{(i_1), \dots, (i_K)\}$ the next properties for $C_K = \frac{H(X_{f_K^*})}{H(X)}$ are posted:

- Deterministic Variable $\Rightarrow C_K \triangleq 1, \forall K \in \{1, \dots, N\}$
- $C_{K+1} \geq C_K, \forall K \in \{1, \dots, N\}$
- $C_0 \triangleq 0$
- $C_N = 1$

In the case of a media with known statistics the estimation of conditional entropy and the consequent *RAMIS* rule can be reduced. The simplest cases correspond to fields with *ID* or *IID* joint probabilities where:

$$H(X_f) = \sum_{i \in f} H(X_i) \stackrel{(\text{Only IID})}{=} |f| \cdot H(X_{i^*}) \quad (\text{C.21})$$

and the resolvability capacity is reduced to:

$$\begin{aligned} C_K &= \frac{\max_{f_K \in \mathbf{F}_K} H(X_{f_K})}{H(X)} \\ &\stackrel{(\text{Only for IID})}{=} \frac{K \cdot H(X_{i^*})}{N \cdot H(X_{i^*})} \\ &= \frac{K}{N} \end{aligned}$$

In real scenarios the field under analysis incorporates some level of spatial interdependence. A case of interest for the presented work is associated with Markov random fields *MRFs* where the spatial dependence can be characterized by Markov chains, while the spatial scope of interaction can be defined by the incorporation of *Cliques*. In simulation of fields with spatial dependence *MRFs* provide an useful tool (Cross and Jain, 1983) as illustrated in figure C.1 but here the focus is in the use of *MRFs* to empirically estimate the spatial dependence in the field.

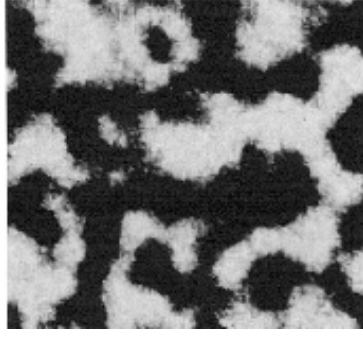


Figure C.1: Synthetic image by *MRF* modeling (Cross and Jain, 1983).

By the use of approximations of *Cliques*, the joint probability models can be described by considering a markovian property. For example, defining $b_{u,v} = x_i$ with $i \in [N]$ and $(u, v) \in [M] \times [M]$ as a mapping transform (from the original stochastic field space without specific spatial order to a sorted 2-D space) and assuming a *Clique* with only vertical and horizontal dependence of unitary distance ($\mathcal{N}(b_{u,v}) = \{b_{u-1,v}, b_{u,v-1}, b_{u+1,v}, b_{u,v+1}\}$), it is possible to reduce the joint probability of a specific spatial variable by only considering the interaction with the members of its *Clique* (only four variables in the neighborhood of this example). A realistic characterization of the *Clique* will provide better estimation for the joint entropy while a simplified version of the *Clique* will offer more computationally affordable implementations. Thus,

$$\mathbb{P}_X(X) = \prod_{(j,k) \in [NxN]} \Psi(b_{j,k}, \mathcal{N}(b_{j,k})), \quad (\text{C.22})$$

where $\Psi(b_{j,k}, \mathcal{N}(b_{j,k}))$, and considering the Joint Entropy as:

$$\tilde{H}(\Psi_{j,k}) = \sum_{b_{j,k} \in \mathcal{A}} \Psi_{j,k} \log(\Psi_{j,k})_{on N} \quad (\text{C.23})$$

implies that: $H(X) = \sum_{l=1}^k \tilde{H}(\Psi_{i_l j_l})$

In the case of resolvability capacity for *MRFs* the target is to study definitions based on

$$\text{Clique's: } C_k = \frac{\max_{f_k \in \mathbf{F}_k} \sum_{l=1}^k \tilde{H}(\Psi_{i_l j_l})}{\sum_{l=1}^{M^2} \tilde{H}(\Psi_{i_l j_l})}$$

C.3.3 Stopping Rule via C_K Evolution

A key topic related with adaptive sampling strategies is to determine a robust method for stopping the sampling process. The objective is to set a limit on the number of samples. For this a stopping criteria has been proposed based on the resolvability capacity of the field.

In particular, the use of the monotonically increasing sequence $\tilde{\mathcal{C}}_K$ has been proposed in the iterative sampling scheme. Using the Eq. (C.19), the change in the resolvability capacity factor, after taking an additional measure, is given by:

$$\begin{aligned}\Delta\tilde{\mathcal{C}}_K &= \tilde{\mathcal{C}}_K - \tilde{\mathcal{C}}_{K-1} \\ &= \frac{I(\tilde{f}_K^*) - I(\tilde{f}_{K-1}^*)}{H(X)} \\ &\propto H(X_{i_K^*} | X_{i_1^*}, \dots, X_{i_{k-1}^*})\end{aligned}\tag{C.24}$$

The use of this proportional version is supported by the fact that the joint entropy for the field X is the same throughout the sensing process, but inaccessible in practical applications. Therefore, $H(X)$ can be considered as a scale factor and, consequently, it could be omitted as shown in Eq. (C.24). For the iterative approach, $\tilde{\mathcal{C}}_K$ is a monotonically decreasing sequence as $K \rightarrow N$. In particular, when $\tilde{\mathcal{C}}_{k_{limit}} = 0$ for a $k_{limit} \in 1, \dots, K$ the remaining non measured $N - k_{limit}$ random variables are deterministically determined by the variables indexed by the rule $\tilde{f}_{k_{limit}}^*$. Then, no more information will be obtained for the random field by additional measures and the sensing process must be truncated.

From monotonically increasing nature of $\tilde{\mathcal{C}}_K$, $\Delta\tilde{\mathcal{C}}_K$ is a monotonically decreasing sequence as $K \rightarrow N$. Finally, when $\Delta\tilde{\mathcal{C}}_k = 0$ after k sequential measures, no more information will be obtained from the field by additional measures and the sensing process must be truncated.

C.4 On Regionalized variables with spatial dependence

Here, it has been worked either without any kind of spatial dependence for random variables or only with very simple assumptions over it dependence for regionalized random variables. The target is to develop and complement a formalization and practical reduction for this kind of dependence.

C.4.1 MRF models and *Cliqu e* structure estimation

When the spatial dependence is headed by a markovian property it is possible to reduce the conditional probabilities by only considering conditionals on the *Cliqu e* of the variable of interest:

$$P(X_i|X^i) \equiv P(X_i|X_{CL(i)}) \quad (C.25)$$

Where $CL(i)$ correspond to the *Cliqu e* associated to the random variable X_i at the position i .

C.4.2 *Cliqu e* estimation and Mutual Information

Given again a random field $X = \{X_i : i \in [N]\}$, the current goal is to define the multi-point conditional probability in terms of two-point conditional probability:

$$P(X_i|X^i) = f(\{P(X_i|X_k); \forall k \in [N] \setminus i\}) \quad (C.26)$$

Using Bayes rule provides the following conditional probabilities:

$$\begin{aligned} P(X_i|X^i) &= \frac{P(X_i, X^i)}{P(X^i)} \\ &= \frac{P(X)}{P(X^i)} \\ &= \frac{P(X^{\{i,k\}}|X_{\{i,k\}})P(X_{\{i,k\}})}{P(X^{\{i,k\}}|X_k)P(X_k)} \\ &= \frac{P(X^{\{i,k\}}|X_{\{i,k\}})}{P(X^{\{i,k\}}|X_k)}P(X_i|X_k) \end{aligned} \quad (C.27)$$

Eq. (C.27) shows that $P(X_i|X^i)$ can be rewritten in terms of $P(X_i|X_k)$ for any k in the complement subset. By using this property it is possible to derive a combination of $P(X_i|X_k)$ that is equal to multi-point conditional probability.

Note that $H(X_i|X_k) + H(X_{\{i,k\}}|X^{\{i,k\}}) - H(X_k|X^{\{i,k\}})$ can be rewritten using mutual information property, $H(X|Y) = H(X) - I(X; Y)$.

$$\begin{aligned}
& H(X_i|X_k) + H(X^{\{i,k\}}|X_{\{i,k\}}) - H(X^{\{i,k\}}|X_k) \\
&= H(X_i) - I(X_i; X_k) + H(X^{\{i,k\}}) - I(X_{\{i,k\}}; X^{\{i,k\}}) - H(X^{\{i,k\}}) + I(X_k; X^{\{i,k\}}) \\
&= H(X_i) - I(X_i; X_k) - I(X_{\{i,k\}}; X^{\{i,k\}}) + I(X_k; X^{\{i,k\}}) \quad (\text{C.31})
\end{aligned}$$

Then, the expression of Eq. (C.30) can be sorted in three kinds of components:

$$\begin{aligned}
I(X_i; X^i) &= H(X_i) \\
&\quad - \frac{1}{|X^i|} \sum_{k \in I_{X^i}} H(X_i) - I(X_i; X_k) - I(X_{\{i,k\}}; X^{\{i,k\}}) + I(X_k; X^{\{i,k\}}) \\
&= \underbrace{\left[\frac{1}{|X^i|} \sum_{k \in I_{X^i}} I(X_i; X_k) \right]}_{\text{one and two-point statistics}} \\
&\quad + \underbrace{\left[\frac{1}{|X^i|} \sum_{k \in I_{X^i}} I(X_{\{i,k\}}; X^{\{i,k\}}) \right]}_{\text{higher order statistics}} - \underbrace{\left[\frac{1}{|X^i|} \sum_{k \in I_{X^i}} I(X_k; X^{\{i,k\}}) \right]}_{\text{redistributable term}} \quad (\text{C.32})
\end{aligned}$$

Multi-point conditional probability can be extended for a lower order subset $X^{k,J}$ where $J \subset [N]^2$ and $k \notin J$.

$$P(x_k|X^{k,J}) = \left(\prod_{l \in I_{X^{k,J}}} \frac{P(X^{k,l,J}|X_{k,l})}{P(X^{k,l,J}|x_l)} P(x_k|x_l) \right)^{\frac{1}{|I_{X^{k,J}}|}} \quad (\text{C.33})$$

This also extends the multi-point mutual information for lower order subset $X^{k,J}$.

$$\begin{aligned}
I(x_k; X^{k,J}) &= \left[\frac{1}{|X^{k,J}|} \sum_{l \in I_{X^{k,J}}} I(x_k; x_l) \right] + \\
&\quad \left[\frac{1}{|X^{k,J}|} \sum_{l \in I_{X^{k,J}}} I(X_{l,k}; X^{l,k,J}) \right] - \left[\frac{1}{|X^{k,J}|} \sum_{l \in I_{X^{k,J}}} I(x_l; X^{l,k,J}) \right] \quad (\text{C.34})
\end{aligned}$$

Using Eq. (C.34), in Eq. (C.32), recursively, it is obtained that:

$$\begin{aligned}
I_{|X|}(x_i; X^i) = & \underbrace{\left[2H(X_i) + \mathcal{H}_{|X|-1} \sum_{\substack{k \in I_X \\ k \neq i}} H(x_k) \right]}_{\text{marginal entropies}} \\
& - \underbrace{\left[\frac{1}{|X|-1} \sum_{k \in I_{X^i}} I(x_i; x_k) + \left(\frac{N}{N-1} - \frac{1}{(N-1)!} \right) \left(\sum_{k \in I_{X^i}} \sum_{j \in I_{X^i, k}} I(x_k; x_j) \right) \right]}_{\text{two-point statistics}} \\
& \underbrace{\left[\sum_{m=1}^{|X|-1} \frac{1}{(m-1)!} \sum_{\substack{s_1 \in I_X \\ s_1 \neq s_0}} \dots \sum_{\substack{s_m \in I_X \\ s_m \neq s_0, \dots, s_{m-1}}} I(X_{s_m, s_{m-1}}; X^{s_0, \dots, s_m}) \right]}_{\text{remaining higher order statistics}} \tag{C.35}
\end{aligned}$$

C.5 Markovian 1D Scenario

A 1D markov chain provides a simple sequence of random variables with spatial dependence. In literature, Markov chains are applied in geology and geostatistics to represent discrete regionalized variables such as lithologies or facies. It is required to note that markovian model do not use variograms or auto-covariance functions in order to measure the spatial structures as most of the available models do. Instead, a markovian model provides a natural way to impose conditional probabilities. It is important, because conditional probabilities could be interpreted geologically much easier than variogram or other mathematical tools.

C.5.1 1-D Markovian Spatial model

The proposed model is the one dimensional array described in Fig. C.2. In this setting, the regionalized variable is composed by N random variables. The array represents N locations with statistical dependence. The final goal is to characterize the conditional statistics of a specific unsampled random variable given a set of sampled locations. In the proposed empirical version of Optimal well placement (the implemented version of *AdSEMES* or *RAMIS*), the conditional statistics are estimated from the number of pattern occurrences in training image.

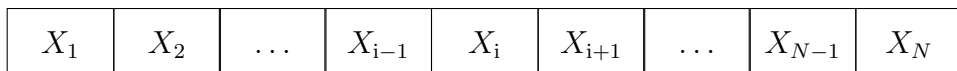


Figure C.2: 1-D regionalized variable.

1-D Markov Chain

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. C.2, a past-present-future sorting has been imposed from location 1 to the location N.

Let $X_1, X_2, \dots, X_{i-1}, X_i, X_{i+1}, \dots, X_{N-1}, X_N$ be a sequence of random variables taking values in a state space ($\mathcal{A} = \{a_1, a_2, \dots, a_{|\mathcal{A}|}\}$). Then, the markovian property states that:

$$Pr(X_i = a_k | X_{i-1} = a_l, X_{i-2} = a_m, \dots, X_1 = a_p) = Pr(X_i = a_k | X_{i-1} = a_l). \quad (\text{C.36})$$

In addition, under the assumption of stationarity and time invariance, the transition probability is given by:

$$p_{l,k} = Pr(X_i = a_k | X_{i-1} = a_l). \quad (\text{C.37})$$

From this transition probability the 1-D markov chain can be described by its transition matrix:

$$p = \begin{pmatrix} p_{1,1} & p_{1,2} & \dots & \dots & p_{1,|\mathcal{A}|} \\ p_{2,1} & & \dots & & \\ \dots & & \dots & & \\ & & \dots & p_{l,k} & \dots \\ \dots & & \dots & & \\ p_{|\mathcal{A}|,1} & & \dots & & p_{|\mathcal{A}|,|\mathcal{A}|} \end{pmatrix} \quad (\text{C.38})$$

with $p_{l,k}$ denoting the probability of transition from a_l to a_k , $\forall l, k \in \{1, \dots, |\mathcal{A}|\}$. This transition probability is called single step transition because is related with the transition between two consecutive elements of the sequence of random variables, $\{X_i\}$. Extending this idea, the transition probability can be characterized relating elements, in the random sequence, separated by M steps by the multiplication of the single step transition matrix by itself M times. Thus:

$$Pr(X_i = a_k | X_{i-M} = a_l, X_{i-M-1} = a_m, \dots, X_1 = a_p) = Pr(X_i = a_k | X_{i-M} = a_l) = (p^M)_{l,k}. \quad (\text{C.39})$$

Joint probability mass of the sequence $\{X_i\}$

For a 1-D markov chain the joint probability mass function can be written as:

$$p(x_1, x_2, \dots, x_N) = p(x_1) \cdot \prod_{r=2}^N p(x_r | x_{r-1}) \quad (\text{C.40})$$

Then, the Joint Entropy for the sequence $\{X_i\}$, using the theorem 2.5.1 in Cover and Thomas (2006), is described by:

$$H(X_1, X_2, \dots, X_N) = \sum_{i=1}^N H(X_i | X_{i-1}, X_{i-2}, \dots, X_1) = \sum_{i=1}^N H(X_i | X_{i-1}) \quad (\text{C.41})$$

The conditional probability can be given from the transition matrix p . For the two point joint probability $p(x_i, x_{i-1})$, using the fact that $p(x_i, x_{i-1}) = p(x_{i-1}) \cdot p(x_i | x_{i-1})$. Finally, given the marginal entropy for the initial state X_1 , the remaining marginal can be expressed by:

$$p(x_i) = \sum_{x_{i-1} \in \mathcal{A}} p(x_{i-1}) \cdot p(x_i | x_{i-1}) \quad (\text{C.42})$$

Conditioning to a subset of states

The main challenge is to apply the proposed *OSP* approach to the *1-D* markov chain. As presented in sec. 3.5, the entropy based *OSP* can be stated in several equivalent ways. First, given a subset of previously measured locations (or at least selected to be measured), the target is to find the non measured location with maximum entropy, $X_i^* = \text{argmax} H(X_i | X_{Measured})$. Alternatively, it can be asked for the location minimizing the posterior entropy for the remaining non measured locations, $X_i^* = \text{argmin} H(X_{NoMeasured} \setminus i | X_{Measured, i})$. These formulations require the joint *pmfs* of the random variables at the unmeasured locations conditioned to the subset of the variables at the measured locations, $p(x_i | x_{Measured})$ or $p(x_{NoMeasured} | x_{Measured})$, respectively.

Conditioning to any arbitrary subset of states

Let $\{X_{b_j}\}$ and $\{X_{a_j}\}$ be two disjoint subsets denoting the past and the futures variables conditioning the state X_i . Then the measured locations can be represented by $X_{Measured} = \{X_{b_j}\} \cup \{X_{a_j}\}$, as illustrated in Fig. C.3.

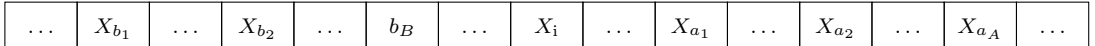


Figure C.3: Measured Variables in separated past and future subsets.

Then, the target conditional entropy for the state X_i , can be defined as:

$$\begin{aligned} H(X_i | \{X_{b_j}\}, \{X_{a_j}\}) \\ = - \sum_{\{x_{b_j}\} \in \mathcal{A}^B} \sum_{x_i \in \mathcal{A}} \sum_{\{x_{a_j}\} \in \mathcal{A}^A} p(\{x_{b_j}\}, x_i, \{x_{a_j}\}) \cdot \log p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) \end{aligned} \quad (\text{C.43})$$

Furthermore, it is possible to rewrite (C.43) by using the definition of the conditional entropy as the expected value of the entropies of the conditional distributions, averaged over the conditioning random variables [Chapter 2.2, Cover and Thomas (2006)].

$$\begin{aligned}
H(X_i | \{X_{b_j}\}, \{X_{a_j}\}) &= \sum_{\{x_{b_j}\} \in \mathcal{A}^B} \sum_{\{x_{a_j}\} \in \mathcal{A}^A} p(\{x_{b_j}\}, \{x_{a_j}\}) \cdot H(X_i | \{X_{b_j}\} = \{x_{b_j}\}, \{X_{a_j}\} = \{x_{a_j}\}) \\
\end{aligned} \tag{C.44}$$

In one hand, to evaluate the expression from (C.43), it is required to work in the conditional *pmf* of the state X_i . Thus, using the relationship of the joint *pmf* and the conditional *pmf*:

$$\begin{aligned}
p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) &= \frac{p(\{x_{b_j}\}, x_i, \{x_{a_j}\})}{p(\{x_{b_j}\}, \{x_{a_j}\})} \\
&= \frac{p(\{x_{b_j}\}, x_i, x_{a_1}, x_{a_2}, \dots, x_{a_A})}{p(\{x_{b_j}\}, x_{a_1}, x_{a_2}, \dots, x_{a_A})} \\
&= \frac{p(x_{a_A} | \{x_{b_j}\}, x_i, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}}) \cdot p(\{x_{b_j}\}, x_i, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}})}{p(x_{a_A} | \{x_{b_j}\}, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}}) \cdot p(\{x_{b_j}\}, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}})} \\
\text{by markovian property} &= \frac{p(x_{a_A} | x_{a_{A-1}}) \cdot p(\{x_{b_j}\}, x_i, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}})}{p(x_{a_A} | x_{a_{A-1}}) \cdot p(\{x_{b_j}\}, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}})} \\
&= \frac{p(\{x_{b_j}\}, x_i, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}})}{p(\{x_{b_j}\}, x_{a_1}, x_{a_2}, \dots, x_{a_{A-1}})} \\
\end{aligned} \tag{C.45}$$

From Eq. (C.45), the effect of the most future state, X_{a_A} , vanishes for the markovian property and the presence of the closer future state, $X_{a_{A-1}}$. Furthermore, iterating the same principle for the rest of the variables in the subset of future states, it can post that:

$$\begin{aligned}
p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) &= \frac{p(\{x_{b_j}\}, x_i, x_{a_1})}{p(\{x_{b_j}\}, x_{a_1})} \\
&= p(x_i | \{x_{b_j}\}, x_{a_1}) \\
\end{aligned} \tag{C.46}$$

From the above result, for a markovian chain, given a subset of future states, x_{a_j} , conditioning the present state, X_i , then the present state depends only the nearest future state that is part of the subset. The essence of this result was presented by Elfeki and Dekking (2001) by the formulation of the conditional probability $p(x_i | x_{i-1}, x_N)$.

In Eq. (C.46), it is required to reduce the dependence related with the subset of past states. Thus,

$$\begin{aligned}
p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) &= \frac{p(\{x_{b_j}\}, x_i, x_{a_1})}{p(\{x_{b_j}\}, x_{a_1})} \\
&= \frac{p(x_{a_1} | \{x_{b_j}\}, x_i) \cdot p(\{x_{b_j}\}, x_i)}{p(x_{a_1} | \{x_{b_j}\}) \cdot p(\{x_{b_j}\})} \\
&\stackrel{\text{by markovian property}}{=} \frac{p(x_{a_1} | x_i) \cdot p(\{x_{b_j}\}, x_i)}{p(x_{a_1} | \{x_{b_j}\}) \cdot p(\{x_{b_j}\})} \\
&= \frac{p(x_{a_1} | x_i) \cdot p(x_i | \{x_{b_j}\}) \cdot p(\{x_{b_j}\})}{p(x_{a_1} | \{x_{b_j}\}) \cdot p(\{x_{b_j}\})}
\end{aligned} \tag{C.47}$$

Then, simplifying and applying the markovian property Eq. (C.47) can be reduced to:

$$\begin{aligned}
p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) &= \frac{p(x_{a_1} | x_i) \cdot p(x_i | \{x_{b_j}\})}{p(x_{a_1} | \{x_{b_j}\})} \\
&\stackrel{\text{by markovian property}}{=} \frac{p(x_{a_1} | x_i) \cdot p(x_i | x_{b_B})}{p(x_{a_1} | x_{b_B})}
\end{aligned} \tag{C.48}$$

In other words,

$$p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) = p(x_i | x_{b_B}, x_{a_1}) \tag{C.49}$$

Summarizing the last results, conditioning a state X_i by any subset of states reduce to conditional bivariate *pmfs* only relating X_i to the nearest past and to the nearest future states inside the set of measured variables.

In the other hand, to calculate the required conditional entropy, the joint *pmf* is needed, $p(\{x_{b_j}\}, x_i, \{x_{a_j}\})$. Applying Eq. (C.40):

$$\begin{aligned}
&p(\{x_{b_j}\}, x_i, \{x_{a_j}\}) \\
&= p(x_{b_1}) \cdot \left[\prod_{h=2}^B p(x_{b_h} | x_{b_{h-1}}) \right] \cdot p(x_i | x_{b_B}) \cdot p(x_{a_1} | x_i) \cdot \left[\prod_{h=2}^A p(x_{a_h} | x_{a_{h-1}}) \right]
\end{aligned} \tag{C.50}$$

In short, Eq. (C.48) and Eq. (C.50) can be expressed in terms of marginal and two points conditional probabilities. At this point, all the information required to efficiently calculate Eq. (C.43) is available, and consequently it is possible to solve the *MES* rule. All conditional probabilities, in Eq. (C.48) and Eq. (C.50), can be estimated from the transition matrix by the n-step transition probabilities from Eq. (C.39). In addition, the marginal probabilities for $p(x_{b_1})$ required in Eq. (C.50), can be iteratively calculated (Cover and Thomas, 2006) by:

$$p(x_i) = \sum_{x_{i-1} \in \mathcal{A}} p(x_{i-1}) \cdot p(x_i | x_{i-1}) \quad (\text{C.51})$$

In brief, as expected for a 1 - D Markov chain, all the conditional entropy characterization relies on the transition matrix and the initial marginal distribution for the state X_1 .

In addition, using Eq. (C.49) the target entropy of the conditional distribution for the state X_i , can be defined as:

$$\begin{aligned} H(X_i | \{X_{b_j}\} = \{x_{b_j}\}, \{X_{a_j}\} = \{x_{a_j}\}) \\ &= - \sum_{x_i \in \mathcal{A}} p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) \cdot \log p(x_i | \{x_{b_j}\}, \{x_{a_j}\}) \\ &\text{by Eq. (C.49)} = - \sum_{x_i \in \mathcal{A}} p(x_i | x_{b_B}, x_{a_1}) \cdot \log p(x_i | x_{b_B}, x_{a_1}) \\ &= H(X_i | X_{b_B} = x_{b_B}, X_{a_1} = x_{a_1}) \end{aligned} \quad (\text{C.52})$$

Then, from Eq. (C.44) and Eq. (C.52), the conditional entropy can be formulated as:

$$H(X_i | \{X_{b_j}\}, \{X_{a_j}\}) = \sum_{\{x_{b_j}\} \in \mathcal{A}^B} \sum_{\{x_{a_j}\} \in \mathcal{A}^A} p(\{x_{b_j}\}, \{x_{a_j}\}) \cdot H(X_i | X_{b_B} = x_{b_B}, X_{a_1} = x_{a_1}) \quad (\text{C.53})$$

A preliminary attempt to reduce the Eq. (C.53) consisted in properly separating the conditional variables:

$$\begin{aligned} &H(X_i | \{X_{b_j}\}, \{X_{a_j}\}) \\ &= \sum_{\{x_{b_j}\} \setminus x_{b_B} \in \mathcal{A}^{B-1}} \sum_{x_{b_B} \in \mathcal{A}} \sum_{\{x_{a_j}\} \setminus x_{a_1} \in \mathcal{A}^{A-1}} \sum_{x_{a_1} \in \mathcal{A}} [\\ &\quad p(\{x_{b_j}\} \setminus x_{b_B}, x_{b_B}, \{x_{a_j}\} \setminus x_{a_1}, x_{a_1}) \cdot H(X_i | X_{b_B} = x_{b_B}, X_{a_1} = x_{a_1})] \\ \text{By Bayes} &= \sum_{\{x_{b_j}\} \setminus x_{b_B} \in \mathcal{A}^{B-1}} \sum_{x_{b_B} \in \mathcal{A}} \sum_{\{x_{a_j}\} \setminus x_{a_1} \in \mathcal{A}^{A-1}} \sum_{x_{a_1} \in \mathcal{A}} [\\ &\quad p(x_{b_B}, x_{a_1} | \{x_{b_j}\} \setminus x_{b_B}, \{x_{a_j}\} \setminus x_{a_1}) \cdot p(x_{b_B}, x_{a_1}) \cdot H(X_i | X_{b_B} = x_{b_B}, X_{a_1} = x_{a_1})] \\ &= H(X_i | X_{b_B}, X_{a_1}) \cdot \sum_{\{x_{b_j}\} \setminus x_{b_B} \in \mathcal{A}^{B-1}} \sum_{x_{b_B} \in \mathcal{A}} \sum_{\{x_{a_j}\} \setminus x_{a_1} \in \mathcal{A}^{A-1}} \sum_{x_{a_1} \in \mathcal{A}} [\\ &\quad p(x_{b_B}, x_{a_1} | \{x_{b_j}\} \setminus x_{b_B}, \{x_{a_j}\} \setminus x_{a_1})] \\ &= H(X_i | X_{b_B}, X_{a_1}) \end{aligned} \quad (\text{C.54})$$

Only for these two conditioning variables:

$$\begin{aligned}
H(X_i | X_{b_B}, X_{a_1}) = & - \sum_{x_{b_B} \in \mathcal{A}} \sum_{x_i \in \mathcal{A}} \sum_{x_{a_1} \in \mathcal{A}} [(\\
& p(X_{b_B} = x_{b_B}) \cdot p(X_i = x_i | X_{b_B} = x_{b_B}) \cdot p(X_{a_1} = x_{a_1} | X_i = x_i) \\
&) \cdot \log p(X_i = x_i | X_{b_B} = x_{b_B}, X_{a_1} = x_{a_1})] \tag{C.55}
\end{aligned}$$

Finally, using Eq. (C.48):

$$\begin{aligned}
H(X_i | X_{b_B}, X_{a_1}) = & - \sum_{x_{b_B} \in \mathcal{A}} \sum_{x_i \in \mathcal{A}} \sum_{x_{a_1} \in \mathcal{A}} [(\\
& p(X_{b_B} = x_{b_B}) \cdot p(X_i = x_i | X_{b_B} = x_{b_B}) \cdot p(X_{a_1} = x_{a_1} | X_i = x_i) \\
&) \cdot \log \left(\frac{p(x_{a_1} | x_i) \cdot p(x_i | x_{b_B})}{p(x_{a_1} | x_{b_B})} \right)] \tag{C.56}
\end{aligned}$$

C.6 Formulation and Implementation of Noisy Sparse Promoting Solvers

Given sampling schemes provided by *OSP*, its comparison with some sparse promoting oriented schemes has been proposed by the modification of classical $L1$ minimizers solver (available in the *L1 magic* and *CVX* software). The contribution has been focused on formulating the signal recovery process as a generalized sampling problem. The idea relies on to take advantage of the sparse nature of channelized structures, the reduced spatial variations of the $2-D$ images in the proposed database and the use of a prior statistical model as an additional source of information related with spatial dependencies and pixel uncertainty.

Therefore, applying principles from *Noisy compressive Sensing (NCS)* to the reconstruction of binary channels of permeability for the estimation of the model, it is possible to use *MPS* realizations as a source of statistical data. In particular, it allows the estimation of variance and covariance of the target regionalized variables.

C.6.1 *NCS* and *Whitening* process

The main idea supporting this preliminary work was to consider the noise associated to measurements in the sparse recovery solver.

The noise associated to measurements has been incorporated in the sparse recovery solver relaxing the search space. In this way, given a small amount of noisy measurements from geologic data (m measures from a field of N variables, with $m \ll N$), the target is to reconstruct the *actual* channel. While traditional *CS* approaches deal with time-invariant sparse signals without error in measurements, the motivation was supported by the next hypothesis: 1) *NCS* provides a theory of signal recovery from highly incomplete noisy information, and 2) *MPS* could provide information about signal variability required to the noise characterization on *NCS* framework.

Thus, a preliminary framework of *NCS* has been implemented and validated oriented to improve *MPS* performance. Standard *CS* implementation for channelized binary structures was proposed and implemented for others members of the *IDS Lab*³. The reader is referred to Calderón et al (2015) for more details on this approximation. Classical conditions and theory of *NCS* was described in sections C.2.1 and C.2.1.

Here, previous works are extended in order to consider noisy measures, providing a framework to noise characterization. In order to apply the theory of *NCS* a signal model with white noise is required, then a whitening process is needed to use the proposed methods. Thus, the next *sensing model* has been considered:

³Information and Decision Systems Laboratory, *IDS Lab*, at Electrical Engineering Department, Universidad de Chile.

$$I = \Phi \cdot Z \quad (\text{C.57})$$

$$X = I(:) \quad (\text{C.58})$$

$$Y_{\text{Sampled}} = A \cdot X + \xi \quad (\text{C.59})$$

with

- Z ($M \times M$) : Signal in transformed domain (in this case a 2-D *DCT* coefficients matrix of size 200 x 200)
- Φ ($M \times M$) : Transform matrix (inverse *DCT* in this work)
- I ($M \times M$) : Image in canonical domain
- X ($N \times N$) : Vectorization of signal in image domain ($N = M \times M$, in this case 40000)
- A ($m \times N$) : Sampling matrix (m vectors randomly taken from $N \times N$ Identity matrix)
- Y_{Sampled} ($m \times 1$) : vector of m measurements (including Hard and Soft Data)
- ξ ($m \times 1$) : vector of noise in measurements, with covariance matrix C_v and mean ξ_{mean} .

At this point the Eq. (C.59) only differs from Calderón et al (2015) approach in the incorporation of noise component. Here, the noise ξ would be a spatially correlated noise, requiring a *whitening* pre-processing.

A signal model with zero mean noise has been described by the subtraction of the mean of noise ξ_{mean} obtaining:

$$Y_{\text{Sampled}} - \xi_{\text{mean}} = A \cdot X + \xi - \xi_{\text{mean}} \quad (\text{C.60})$$

Defining zero mean variables and rewriting the expression in Eq. (C.61) is:

$$Y_0 = A \cdot X + \xi_0 \quad (\text{C.61})$$

From Eq. (C.61) it is required an additional process to obtain a model with a non correlated noise. In order to achieve a sensing model under white noise and assuming the existence of an invertible covariance matrix for ξ_0 the next formulation has been achieved:

$$C_v^{-\frac{1}{2}} \cdot Y_0 = C_v^{-\frac{1}{2}} \cdot A \cdot X + C_v^{-\frac{1}{2}} \cdot \xi_0 \quad (\text{C.62})$$

$$C_v^{-\frac{1}{2}} \cdot Y_0 = C_v^{-\frac{1}{2}} \cdot A \cdot X + \eta \quad (\text{C.63})$$

$$\hat{Y} = C_v^{-\frac{1}{2}} \cdot A \cdot X + \eta \quad (\text{C.64})$$

Finally, replacing the vectorization process:

$$\hat{Y} = C_v^{-\frac{1}{2}} \cdot A \cdot VEC(\Phi \cdot Z) + \eta \quad (\text{C.65})$$

The Eq. (C.65) fits the classical framework of *NCS* for signals under white noise model. The selection of the sampling matrix A satisfies isotropic property, the vectorization process $VEC(\cdot)$ retain spatial dependence in the regionalized field, the basis DCT provides a domain where the signal is compressible, and C_v is estimated as the experimental covariance of realizations of *MPS*.

C.6.2 On Noisy Compressive Sensing Performance

NCS theory has been a motivation to study the spatial co-dependencies of the regionalized variable of interest. heretofore, how *MPS* can help in achieving this task have been investigated and it generated a preliminary analysis of how the incorporation of spatial dependence improves sparse promoting algorithms.

Statistical Analysis from *MPS*

Using the data base of multichannel MC_1 model, obtained 200 *MPS* realizations has been obtained from several sampling rates. For each sampling regime, the second order statistics has been estimated. As shown in fig. C.4 the field variance in low rate sampling regimes is higher than in regimes with more measurements. In the extreme case of 0.1% measurements, the variance at most field positions was extremely uncertain. As the measurements increased a reduction of non local variance has been observed, thereby the powerful of spatial conditioning from *MPS* has been validated.

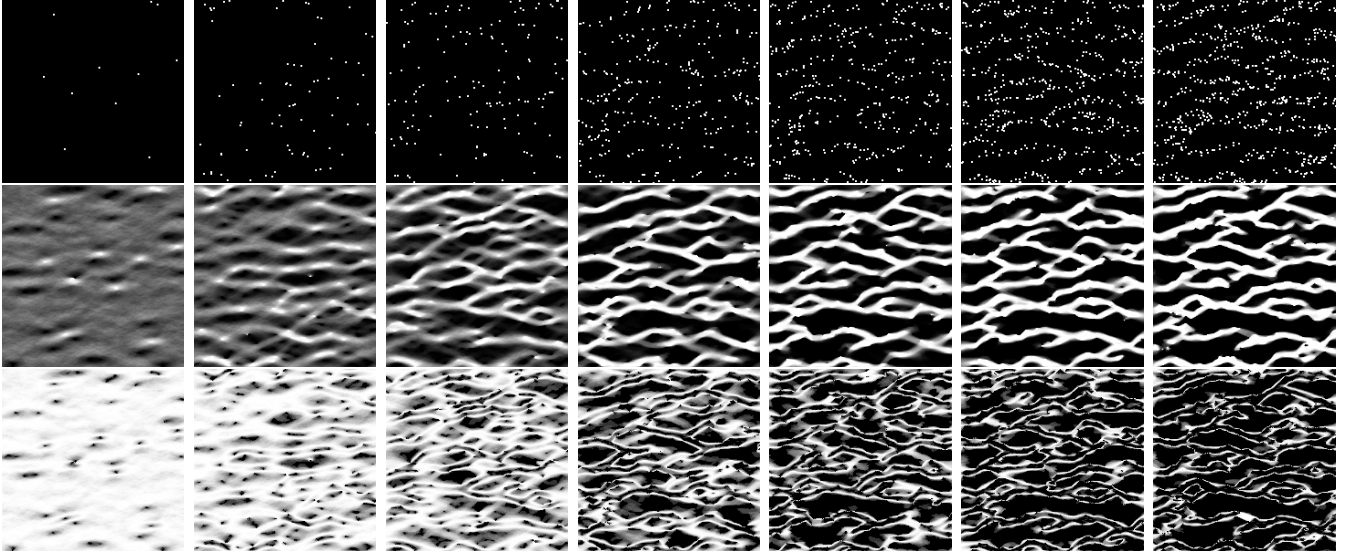


Figure C.4: Statistics from simulations.

For scenarios considering 0.1 %, 0.5 %, 1 %, 2 %, 3 %, 4 %, and 5 % of hard data measures.

As expected for *MPS*, these preliminary outcomes shown that statistical analysis from *MPS* realizations provides some kind of information about the variability of the value of a non measured pixel from the knowledge of positions and values of hard data measurements. This validate the use of *MPS* realizations as an estimation of the spatial covariance between regionalized variables conforming the stochastic field.

In next subsections several attempts of incorporating information from *MPS* realizations in the *NCS* problem has been explored. The main idea was to provide an estimation of covariance matrix and to use it in whitening process.

***NCS*. Naive Approach I. what Covariance?**

The most simple assumption was considered that there was no correlation between regionalized variables, and that all variables was modeled incorporating only with noise with zero mean and the same variance for all the regionalized variables. Under this assumption the variance noise at each individual variable was estimated directly by pixel variance estimation from *MPS* realizations.

At standard *CS* approaches only measurements are considered as input to the reconstruction methods and no one knowledge is given for unmeasured regionalized variables. The incorporation of *MPS* allows not only statistical analysis for global stochastic field, but also the opportunity of considering simulated regionalized variables as noisy virtual measures.

Here most naive *NCS* approach correspond to consider the model from Eq. (C.65) but with C_v^{-1} equal to the identity matrix of proper size. Thus, using a mixture of hard data measures and simulated *virtual* measures in standard *CS* does not take advantage of the

statistical information provided for the proposed *MPS* analysis because all measurements are considered without uncertainty. The above entails that the *CS* reconstruction would be more close to the specific *MPS* realization than to the real target image.

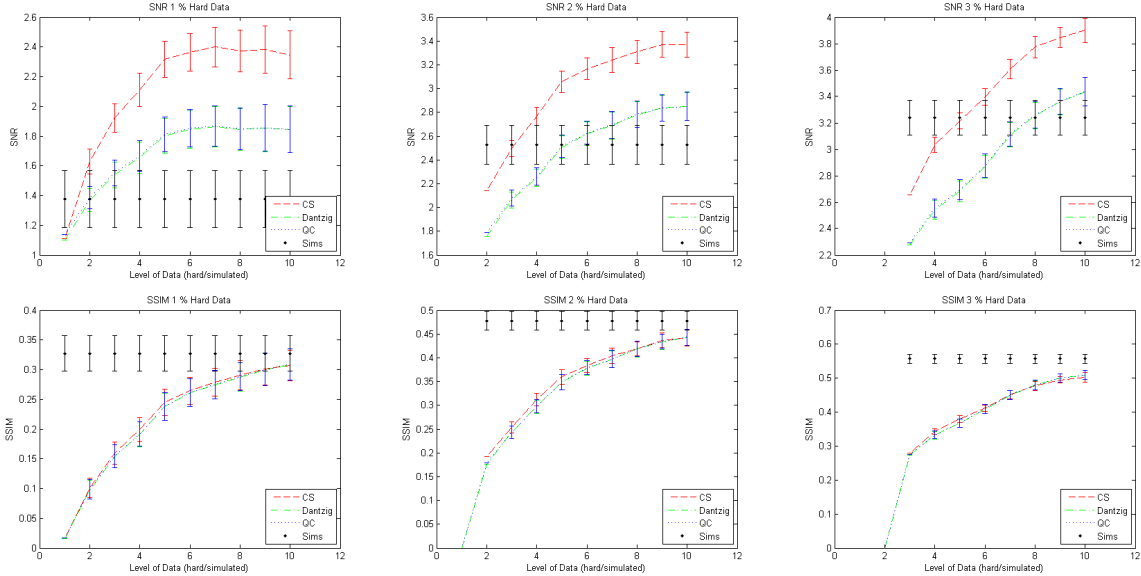


Figure C.5: Performance Analysis of Naive Approach of *NCS* by only relaxing restrictions on soft data.

From an initial amount of hard data measurements some level of soft virtual measurements are added from simulations and then *CS* and naive *NCS* approaches are applied. Upper row *SNR* analysis, lower row *SSIM* analysis. From left to right : initial 1% of hard data , initial 2% of hard data, and initial 3% of hard data.

As shown in fig. C.5, partial results on the most naive *NCS* approach present a lower performance than classical *CS*. As reference the performance of *MPS* is shown by the statistics of simulations (mean and variance), *CS* curves consider hard data and soft data as fixed measures while *NCS* allows some uncertainty for these measures. For *SSIM* indicator all reconstructions curves has very close performance, while for *SNR* indicator the classical *CS* achieves a better global performance.

Without incorporating spatial correlations on the regionalized variables no advantage from *NCS* theory has been achieved. Including uncertainty on measures only increase the searching space for the optimization algorithms. Thus, the next motivation was to include additional constraints promoting desired features on the regionalized variables.

The next step has corresponded to incorporate proposed amendments from section C.2.1. Using alternate recovery models promoting sparsity on signal gradients instead on the signal itself.

In figure C.6, several configurations of hard and virtual data are presented but considering reconstruction approaches using total variation. In terms of visual inspection is possible to appreciate an improvement in the performance of *NCS*. This behavior is confirmed by analysis of the metrics *SNR* and *SSIM* as shown in fig. C.7.

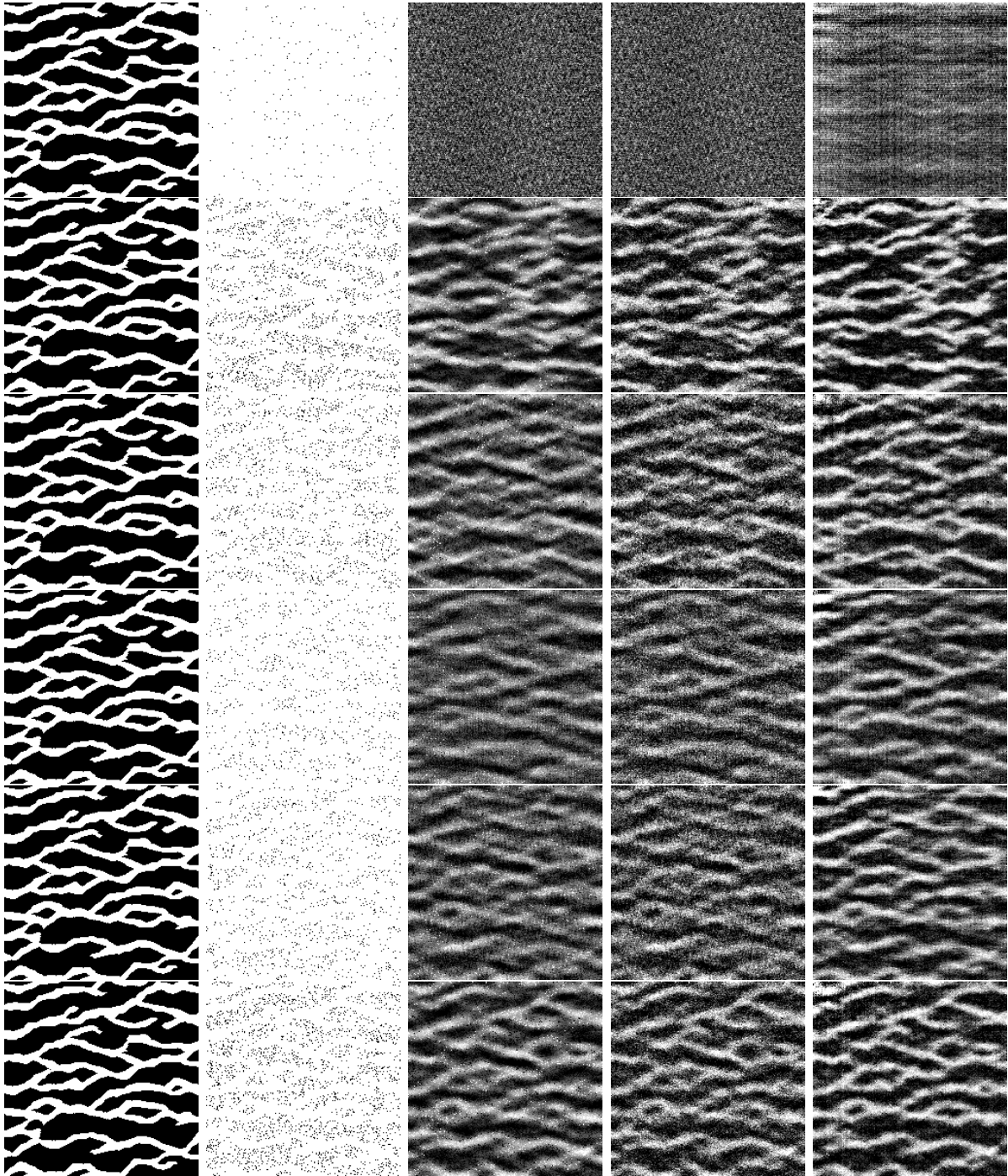


Figure C.6: Examples of outcomes for *NCS* without considering spatial dependence.

From left to right: Target field image, hard data plus simulated data (HD+SD level), standard *CS* reconstruction for HD+SD, *NCS* for HD+SD by Dantzig selector approach, *NCS* for HD+SD by Dantzig selector and *TV* approach. Rows: Different levels of hard data and simulated data from *MPS* realizations: 1 % *HD* plus 0 % *SD*, 1 % *HD* plus 9 % *SD*, 2 % *HD* plus 8 % *SD*, 4 % *HD* plus 6 % *SD*, 5 % *HD* plus 5 % *SD*, 5 % *HD* plus 10 % *SD*

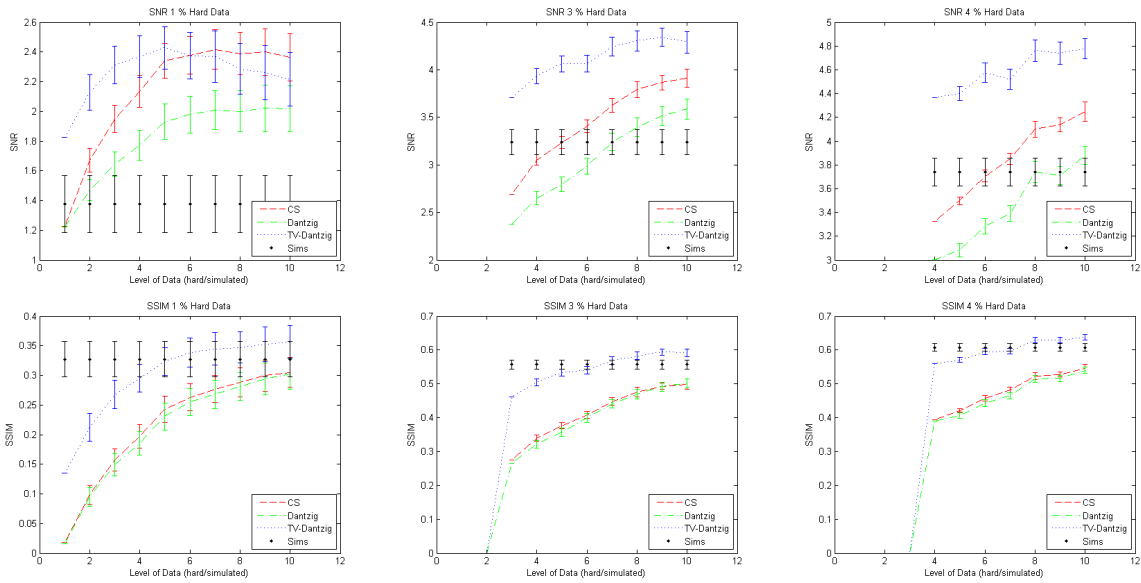


Figure C.7: Performance Analysis of Naive Approach of *NCS* by only relaxing restrictions on soft data adding *TV* as sparsity promoting approach.

From an initial amount of hard data measurements some level of soft virtual measurements are added from simulations and then *CS* and naive *NCS* approaches are applied. Upper row *SNR* analysis, lower row *SSIM* analysis. From left to right : initial 1% of hard data , initial 3% of hard data, and initial 4% of hard data.

NCS. Naive Approach II. Spatial Independence

Here, regionalized variables are considered independent but for each pixel the noise variance is estimated from the empirical variance by the values simulated at this pixel. Thus far only qualitative results has been obtained at this point which account for an apparent improvement in the method's ability to capture something of the structure of original image (figs. C.8, C.9, C.10 and C.11).

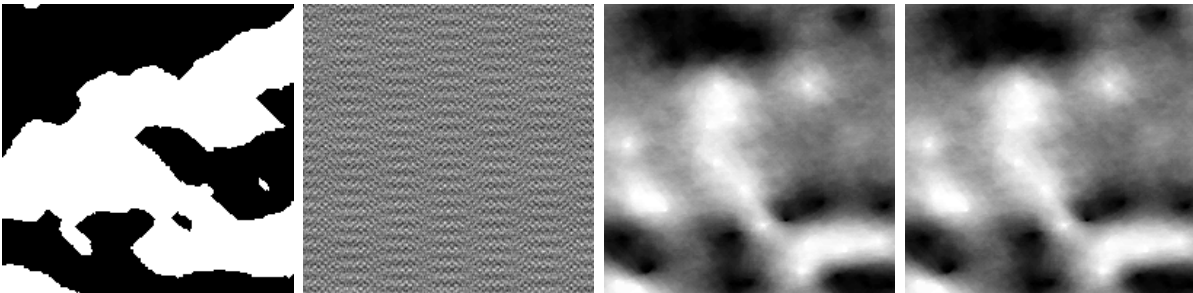


Figure C.8: Example of *NCS* reconstruction under assumption of independence: SC1. Harddata 0.1 %. True Image. Standard *CS*. *NCS* by quadratic constraints. *NCS* by Dantzig selector.

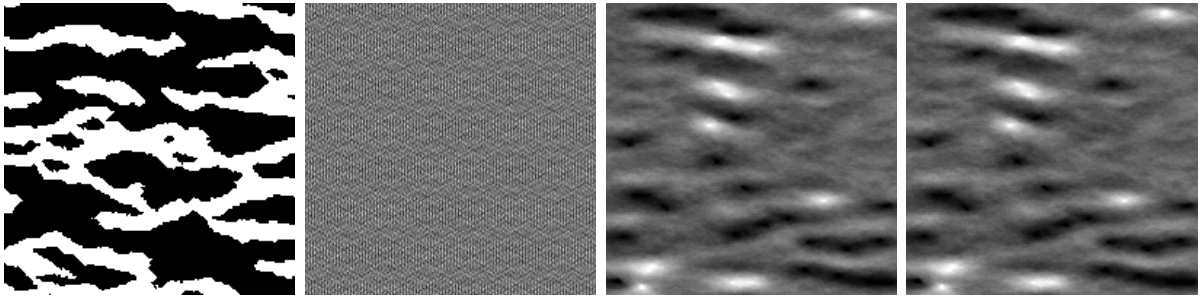


Figure C.9: Example of *NCS* reconstruction under assumption of independence: MC1.
 Harddata 0.1 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig.

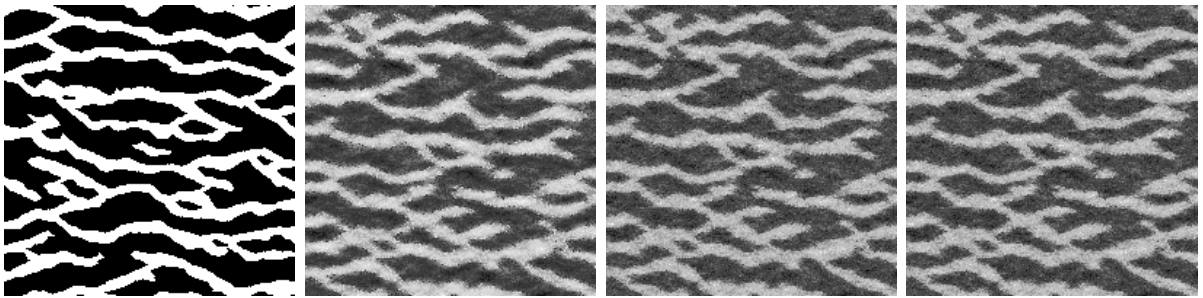


Figure C.10: Example 1 of *NCS* reconstruction under assumption of independence: MC2.
 Harddata 0.1 %, Soft data 29.9 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.

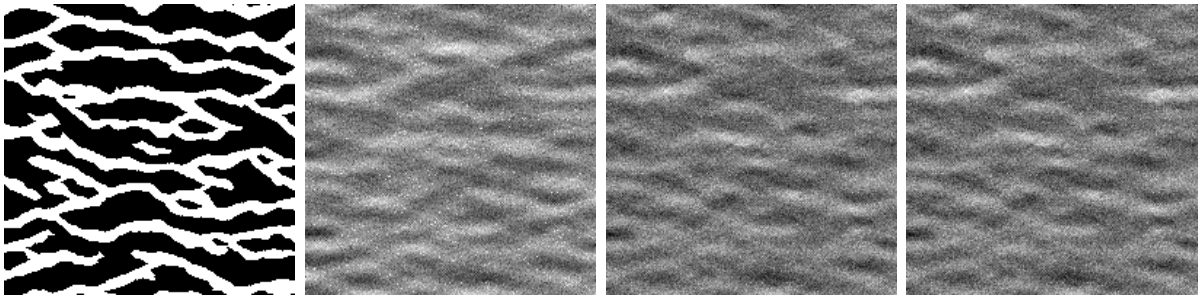


Figure C.11: Example 2 of *NCS* reconstruction under assumption of independence: MC2
 Harddata 0.1 %, Soft data 3.9 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.

NCS. Approach III. Full Covariance

Here, regionalized variables are considered with its full spatial dependence. The spatial dependence requires a dense covariance matrix of size N by N demanding more computational resources. Thus far only qualitative results has been obtained which account for an apparent improvement in the method's ability to capture both global structure and details of original image from low sampling regimes (figs. C.12 and C.13).

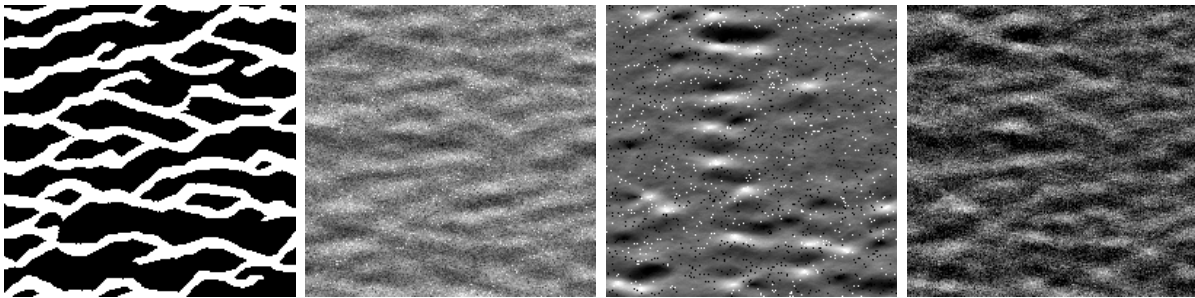


Figure C.12: Example 1 of *NCS* with full covariance estimation: MC2.
 Harddata 0.1 %, Soft data 3.9 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.

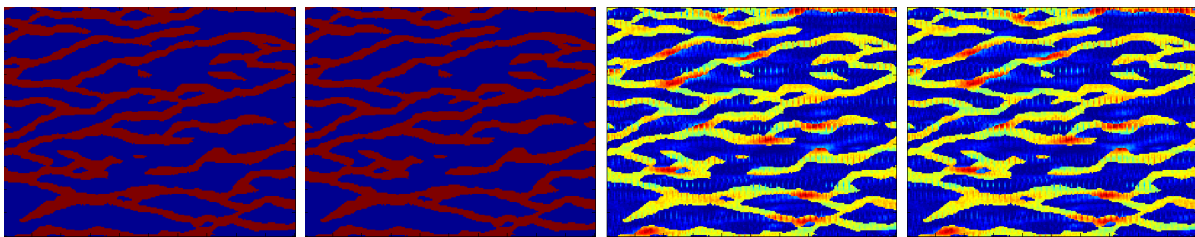


Figure C.13: Example 2 of *NCS* with full covariance estimation: MC2.
 Harddata 1 %, Soft data 99 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.

Bibliography

- Abellan A, Noetinger B (2010) Optimizing subsurface field data acquisition using information theory. *Mathematical Geosciences* 42(6):603–630, doi: 10.1007/s11004-010-9285-6, URL <https://doi.org/10.1007/s11004-010-9285-6>
- Abzalov MZ, Menzel B, Wlasenko M, Phillips J (2010) Optimisation of the grade control procedures at the yandi iron-ore mine, western australia: geostatistical approach. *Applied Earth Science* 119(3):132–142, doi: 10.1179/1743275811Y.0000000007
- Afshari S, Pishvaie M, Aminshahidy B (2014) Well placement optimization using a particle swarm optimization algorithm, a novel approach. *Petroleum Science and Technology* 32(2):170–179
- Arpat B, Caers J (2007a) Conditional simulations with patterns. *Math Geol* 39(2):177–203
- Arpat GB, Caers J (2007b) Conditional simulation with patterns. *Mathematical Geology* 39(2):177–203, doi: 10.1007/s11004-006-9075-3
- Aspie D, Barnes RJ (1990) Infill-sampling design and the cost of classification errors. *Mathematical Geology* 22(8):915–932
- Bangerth W, Klie H, Matossian V, Parashar M, Wheeler M (2005) An autonomic reservoir framework for the stochastic optimization of well placement. *Cluster Computing* 8:255–269
- Bangerth W, Klie H, Wheeler MF, Stoffa P, Sen M (2006) On optimization algorithms for the reservoir oil well placement problem. *Comp Geosciences* 10:303–319
- Baraniuk RG, Davenport M, DeVore R, Wakin M (2008) A simple proof of the restricted isometry property for random matrices. *Constructive Approximation* 28(3):253–263
- Bittencourt AC, Horne RN (1997) Reservoir development and design optimization. In: *SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, San Antonio, Texas, SPE 38895*, pp 1–14
- Blackwell G (1998) Relative kriging errors - a basis for mineral resource classification. *Exploration Mining Geology* 7:99–105
- Boisvert JB, Leuangthong O, Ortiz JM, Deutsch CV (2008) A methodology to construct training images for vein type deposits. *Computers & Geosciences* 34(5):491–502
- Boucheron, Stéphane, Bousquet, Olivier, Lugosi, Gábor (2005) Theory of classification: a survey of some recent advances. *ESAIM: PS* 9:323–375, doi: 10.1051/ps:2005018, URL <https://doi.org/10.1051/ps:2005018>
- Boyko N, Karamemis G, Kuzmenko V, Uryasev S (2011) Sparse Signal Reconstruction: LASSO and Cardinality Approaches. *iseuffedu* URL <http://www.ise.ufl.edu/>

uryasev/files/2012/02/tsparse_2_1_Stan2_Gulver_2.pdf

- Boyko N, Karamemis G, Kuzmenko V, Uryasev S (2014) Sparse Signal Reconstruction: LASSO and Cardinality Approaches, Springer International Publishing, Cham, pp 77–90
- Brus DJ, Heuvelink GBM (2007) Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 138:86–95
- Bui H, La C, Do M (2015) A fast tree-based algorithm for compressed sensing with sparse-tree prior. *Signal Processing* 108(Complete):628–641, doi: 10.1016/j.sigpro.2014.10.026
- Calderón H, Santibanez F, Silva JF, Ortiz JM, Egaña A (2016) Channelized facies recovery based on weighted compressed sensing. In: 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Rio de Janeiro, Brazil, July 10-13, 2016, pp 1–5, doi: 10.1109/SAM.2016.7569627, URL <https://doi.org/10.1109/SAM.2016.7569627>
- Calderon* H, **Santibañez* † F**, Silva JF, Ortiz JM, Egaña A (2019) Geological facies recovery based on weighted ℓ_1 -regularization. *Mathematical Geosciences* online first:1–46
- Calderón H, Silva JF, Ortiz JM, Egaña A (2015) Reconstruction of channelized geological facies based on {RIPless} compressed sensing. *Computers & Geosciences* 77:54 – 65, doi: <http://dx.doi.org/10.1016/j.cageo.2015.01.006>, URL <http://www.sciencedirect.com/science/article/pii/S0098300415000072>
- Candes E, Tao T (2007) The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics* 35(6):2313–2351, doi: 10.1214/009053606000001523, URL <http://www-stat.stanford.edu/~candes/papers/DantzigSelector.pdf>
- Candes E, Romberg J, Tao T (2006a) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on* 52(2):489–509, doi: 10.1109/TIT.2005.862083
- Candes EJ (2008) The restricted isometry property and its applications for compressed sensing. *C R Acad Sci Paris I* 346:589–592
- Candes EJ, Romberg J, Tao T (2006b) Robust uncertainty principle: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* 52(2):489–509
- Candes EJ, Romberg J, Tao T (2006c) Stable signal recovery from incomplete and inaccurate measurements. *Commun on Pure and Appl Math* 59:1207–1223
- Chambolle A, Caselles V (2010) An introduction to total variation for image analysis. *Theoretical Foundations and Numerical Methods for Sparse Recovery* 9:263 – 340, URL <http://books.google.com/books?hl=en&lr=&id=MK1E-dch80MC&oi=fnd&>

pg=PA263&dq=An+introduction+to+Total+Variation+for+Image+Analysis&ots=16tLYs3fsS&sig=XJGcUhggS16FDDtGbfvfhifKyw

- Chilés JP, Delfiner P (1999) Geostatistics : modeling spatial uncertainty. Wiley series in probability and statistics, Wiley, New York, URL <http://opac.inria.fr/record=b1098313>, a Wiley-Interscience publication.
- Chiles JP, Delfiner P (2012) Geostatistics: Modeling Spatial Uncertainty, 2nd edn. Willey
- Christakos G, Killam BR (1993) Sampling design for classifying contaminant level using annealing search algorithms. *Water Resources Research* 29(12):4063–4076
- Christodoulou S, Gagatsis A, Xanthos S, Kranioti S, Agathokleous A, Fragiadakis M (2013) Entropy-based sensor placement optimization for waterloss detection in water distribution networks. *Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA)* 27(13):4443–4468, URL <https://EconPapers.repec.org/RePEc:spr:waterr:v:27:y:2013:i:13:p:4443-4468>
- Cohen A, Dahmen W, DeVore R (2009) Compressed sensing and best k -term approximation. *Journal of the American Mathematical Society* 22(1):211–231
- Cover TM, Thomas JA (2006) Elements of Information Theory, 2nd edn. Wiley Interscience, New York
- Cressie N, Gotway C, Grondona M (1990) Spatial prediction for networks. *Tech. Rep. 7:251-271*, Chermometr Intell Lab. Syst
- Cross GR, Jain AK (1983) Markov random field texture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-5(1):25–39*, doi: 10.1109/TPAMI.1983.4767341
- Davis DT, Zhengxiao Chen, Jenq-Neng Hwang, Tsang L, Njoku E (1995) Solving inverse problems by bayesian iterative inversion of a forward model with applications to parameter mapping using smmr remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing* 33(5):1182–1193, doi: 10.1109/36.469482
- Deutsch CV, Journel AG (1998) GSLIB - geostatistical software library and user’s guide. Oxford University Press URL <https://searchworks.stanford.edu/view/2458695>
- Deutsch CV, Magri E, Norrena K (2000) Optimal grade control using geostatistics and economics: methodology and examples. *SME Transactions* 308:43–52
- Dimitrakopoulos R, Godoy M (2014) Grade control based on economic ore/waste classification functions and stochastic simulations: examples, comparisons and applications. *Mining Technology* 123(2):90–106, doi: 10.1179/1743286314Y.0000000062
- Donoho DL (2006) Compressed sensing. *IEEE Transactions on Information Theory*

- Elad M (2010) *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st edn. Springer Publishing Company, Incorporated
- Eldar YC (2015) *Sampling Theory: Beyond Bandlimited Systems*, 1st edn. Cambridge University Press, New York, NY, USA
- Eldar YC, Kutyniok G (eds) (2012) *Compressed sensing : theory and applications*. Cambridge University Press, Cambridge, New York, URL <http://opac.inria.fr/record=b1134461>
- Elfeki A, Dekking M (2001) A markov chain model for subsurface characterization: Theory and applications. *Mathematical Geology* 33(5):569–589, doi: 10.1023/A:1011044812133, URL <http://dx.doi.org/10.1023/A:1011044812133>
- Foucart S, Lai M (2009) Sparsest solutions of underdetermined linear systems via ℓ_p -minimization. *Appl Comput Harmon Anal* 26:395–407
- Francois-Bongarcon D (1983) The practice of sampling of broken ores. *CIM Bulletin* 86(970):75–81
- Gao H, Wang J, Zhao P (1996) The updated kriging variance and optimal sample design. *Mathematical Geology* 28(3):295–313
- Goodchild F, Buttenfield B, Wood J (1994) In *Visualization in Geographical Information Systems*, John Wiley & Sons, chap Introduction to visualizing data validity, pp 141–149
- Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. *Geoderma* 103:3–26
- Gray R, Davisson LD (2004) *Introduction to Statistical Signal Processing*. Cambridge Univ Press
- van Groenigen J, Siderius W, Stein A (1999) Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 87:239–259
- Guardiano F, Srivastava M (1993) Multivariate geostatistics: Beyond bivariate methods. In: *Geostatistics-Troia*, Kluwer Academic, Amsterdam, pp 133–144
- Guestrin C, Krause A, Singh A (2005) Near-optimal sensor placements in Gaussian processes. In: *International Conference on Machine Learning (ICML)*
- Gutjahr A (1991) Geostatistics for sampling designs and analysis. In: Nash Rea (ed) *Groundwater Residue Sampling Design*, American Chemical Society, ACS Symposium Series, Washington, DC, pp 48–90
- Güyagüler Baris (2002) *Optimization of Well Placement and Assessment of Uncertainty*. Doctor of philosophy thesis, Stanford University

- Handwerger D, Ma YZ, Sodergren T (2016) Chapter 11 - effective core sampling for improved calibration of logs and seismic data. In: Holditch Y, Zee MA (eds) *Unconventional Oil and Gas Resources Handbook*, Gulf Professional Publishing, Boston, pp 335 – 360, doi: <http://dx.doi.org/10.1016/B978-0-12-802238-2.00011-0>, URL <http://www.sciencedirect.com/science/article/pii/B9780128022382000110>
- Hassanipak AA, Sharafodin M (2004) Get: A function for preferential site selection of additional borehole drilling. *Explor Mining Geol* 13(1-4):139–146
- Holden L, Hauge R, Skare O, Skorstad A (1998) Modeling of fluvial reservoirs with object models. *Mathematical Geology* 30(5):473–496, doi: 10.1023/A:1021769526425
- Huang T, Lu DT, Li X, Wang L (2013) Gpu-based snesim implementation for multiple-point statistical simulation. *Comput Geosci* 54:75–87, doi: 10.1016/j.cageo.2012.11.022, URL <http://dx.doi.org/10.1016/j.cageo.2012.11.022>
- Journel A, Kyriakidis PC (2004) *Evaluation of Mineral Reserves - A Simulation Approach*. Oxford University Press
- Journel AG (1974) Geostatistics for conditional simulation of orebodies. *Economic Geology* 69:673–687
- Kay SM (1993) *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA
- Kennedy BA (1990) *Surface Mining*, 2nd edn. Society of Mining, Metallurgy and Exploration, Inc.
- Kitanidis PK (1997) *Introduction to geostatistics: applications to hydrogeology*. Cambridge Univ Press
- Kochenderfer MJ, Amato C, Chowdhary G, How JP, Reynolds HJD, Thornton JR, Torres-Carrasquillo PA, Ure NK, Vian J (2015) *Decision Making Under Uncertainty*. MITP
- Krause A (2008) *Optimizing sensing: Theory and applications*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, aAI3343461
- Krause A, Guestrin C, Gupta A, Kleinberg J (2006) Near-optimal sensor placements: Maximizing information while minimizing communication cost. In: *Proc. of Information Processing in Sensor Networks (IPSN)*
- Krause A, Leskovec J, Guestrin C, VanBriesen J, Faloutsos C (2008a) Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management* 134(6):516–526
- Krause A, Singh A, Guestrin C (2008b) Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research (JMLR)* 9:235–284

- Krause A, Rajagopal R, Gupta A, Guestrin C (2009) Simultaneous placement and scheduling of sensors. In: Proc. ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)
- Krause A, Guestrin C, Gupta A, Kleinberg J (2011) Robust sensor placements at informative and communication-efficient locations. *ACM Transactions on Sensor Networks (TOSN)* 7(4), doi: 10.1145/1921621.1921625
- Lam NSN (1983) Spatial interpolation methods: A review. *The American Cartographer* 10(2):129–150, doi: 10.1559/152304083783914958, URL <http://www.tandfonline.com/doi/abs/10.1559/152304083783914958>, <http://www.tandfonline.com/doi/pdf/10.1559/152304083783914958>
- Lloyd C, Atkinson P (1998) Scale and the spatial structure of landform: optimising sampling strategies with geostatistics, URL <https://eprints.soton.ac.uk/17634/>, first published on CD-ROM. Produced by: R.J.Abrahart. Publisher: "GeoComputation CD-ROM"
- MacKay DJC (2002) *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA
- Magnant Z (2011) Numerical methods for optimal experimental design of ill-posed problems. PhD thesis, Emory University, URL <https://search.proquest.com/docview/881634811?accountid=14621>
- Magri E, Ortiz J (2000) Estimation of economic losses due to poor blast hole sampling in open pits. In: *Geostatistics 2000, Proceedings of the 6th International Geostatistics Congress, Cape Town, South Africa, vol 2*, pp 732–741
- Magri E, Ortiz J, Libano R (2010) The economic optimisation of advanced drilling grids for short term planning and grade control at El Tesoro copper mine. In: *MININ 2010, Proceedings of the 4th International Conference on Mining Innovation, Santiago, Chile*, pp 89–98
- Magri E, Ortiz J, Moya R, Salazar A (2011) Advanced reverse circulation drilling as a replacement to blasthole sampling: increasing short term planning profitability at cerro colorado copper mine. In: *5th World Conference on Sampling and Blending, Santiago, Chile*
- Man HQ, Jarzyna A (2013) Integration of core, well logging and 2d seismic data to improve a reservoir rock model : a case study of gas accumulation in the ne polish carpathian foredeep. *Geological Quarterly Vol. 57, No. 2*:289 – 306
- Marchant B, Lark R (2007) Optimized sample scheme for geostatistics surveys. *Math Geol* 39:113–134
- Mariethoz G, Caers J (2014) *Multiple-point Geostatistics: Stochastic Modeling with Training Images*. Wiley, URL <https://books.google.cl/books?id=kQX2oAEACAAJ>

- Matheron G (1971) The Theory of Regionalized Variables and its Applications. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Ecole des Mines de Paris
- McBratney A, Webster R (1981) The design of optimal sampling schemes for local estimation and mapping of regionalized variables. ii. program and examples. *Computers and Geosciences* 7:335–365
- McBratney A, Webster R, Burgess T (1981) The design of optimal sampling schemes for local estimation and mapping of regionalized variables. i. theory and method. *Computers and Geosciences* 7:331–334
- Moon TK, Stirling WC (2007) *Mathematical Methods and Algorithms for Signal Processing*. Pearson Education, New Delhi
- Norrena KP, Deutsch CV (2002) Automatic determination of well placement subject to geostatistical and economic constraints. In: *SPE International Thermal Operations and Heavy Oil Symposium and International Horizontal Well Technology Conference*, Society of Petroleum Engineers, Calgary, AB, Canada, SPE 78996, pp 1–12
- Norris J (1997) *Markov Chains*. Cambridge Univ Press
- Olea RA (1984) Sampling design optimization for spatial functions. *Mathematical Geology* 16(4):369–392
- Oliver DS, Reynolds AC, Liu N (2008) *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge University Press, Cambridge, doi: 10.1017/CBO9780511535642, URL http://books.google.com/books?hl=en&lr=&id=oPuy30fbtFIC&oi=fnd&pg=PR11&dq=Inverse+Theory+for+Petroleum+Reservoir+Charaterization+and+History+Matching&ots=9fIWq86eCn&sig=nrnDTbgY36_iqIEemjzFNw9Xu54http://ebooks.cambridge.org/ref/id/CBO9780511535642
- Onwunalu J, Durlofsky L (2009) Application of a particle swarm optimization algorithm for determining optimum well location and type. *Computational Geosciences* 14(1):183–198, doi: 10.1007/s10596-009-9142-1, URL <http://dx.doi.org/10.1007/s10596-009-9142-1>
- Ortiz JM (2003) Characterization of high order correlation for enhanced indicator simulation, Unpublished doctoral dissertation. PhD thesis, University of Alberta
- Ortiz JM, Deutsch CV (2004) Indicator simulation accounting for multiple-point statistics. *Mathematical Geology* 36(5):545–565, doi: 10.1023/B:MATG.0000037736.00489.b5
- Ortiz JM, Magri EJ (2014) Designing an advanced RC drilling grid for short-term planning in open pits: three case studies. *Journal of the Southern African Institute of Mining and Metallurgy* 114:631–637

- Ortiz JM, Magri EJ, R L (2012) Improving financial returns from mining through geostatistical simulation and the optimized advance drilling grid at el tesoro copper mine. *Journal of the Southern African Institute of Mining and Metallurgy* 112(1):15–22
- Ostroumov V, Rachold V, Vasiliev A, Sorokovikov V (2005) An application of a markov-chain model of shore erosion for describing the dynamics of sediment flux. *Geo-Marine Letters* 25(2):196–203, doi: 10.1007/s00367-004-0201-2, URL <http://dx.doi.org/10.1007/s00367-004-0201-2>
- Peschel GJ, Mokošch M (1991) Interrelations between geostatistics and information theory and their practical use. *Mathematical Geology* 23(1):3–7, doi: 10.1007/BF02065960, URL <https://doi.org/10.1007/BF02065960>
- Pitard F (1993) *Pierre Gy's Sampling Theory and Sampling Practice - Heterogeneity, Sampling Correctness and Statistical Process Control*, 2nd edn. CRC Press, Boca Raton, FL
- Proakis JG, Manolakis DG (1996) *Digital Signal Processing (3rd Ed.): Principles, Algorithms, and Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA
- Raef A, Mattern F, Philip C, Totten M (2015) "3d seismic attributes and well-log facies analysis for prospect identification and evaluation: Interpreted palaeoshoreline implications, weirman field, kansas, {USA} ". *Journal of Petroleum Science and Engineering* 133:40 – 51, doi: <http://dx.doi.org/10.1016/j.petrol.2015.04.028>, URL <http://www.sciencedirect.com/science/article/pii/S0920410515001710>
- Remy N, Boucher A, Wu J (2009) *Applied geostatistics with SGeMS : a user's guide*. Cambridge : Cambridge University Press, formerly CIP
- Rossi ME, Deutsch CV (2014) *Mineral Resource Estimation*. Springer
- Santamarina JC, Fratta D (2005) *Discrete signals and inverse problems : an introduction for engineers and scientists*. J. Wiley and sons, Chichester, URL <http://opac.inria.fr/record=b1105865>
- Scheidt C, Caers J (2009) Representing spatial uncertainty using distances and kernels. *Mathematical Geosciences* 41(4):397–419, doi: 10.1007/s11004-008-9186-0
- Schweizer D, Blum P, Butscher C (2017) Uncertainty assessment in 3-d geological models of increasing complexity. *Solid Earth* 8(2):515–530, doi: 10.5194/se-8-515-2017, URL <https://www.solid-earth.net/8/515/2017/>
- Shannon CE (1948) A mathematical theory of communication. *Bell System Tech, J* 27:379–423; 623–656
- Silva Maureira DA (2015) *Enhanced geologic modeling with data-driven training images for improved resources and recoverable reserves*, doctoral dissertation. PhD thesis, University of Alberta

- Starck JL, Murtagh F, Fadili J (2010) *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, New York, NY, USA
- Strebelle S (2002) Conditional simulation of complex geological structures using multiple points statistics. *Math Geol* 34(1):1–22
- Strebelle S, Zhang T (2004) Non-stationary multiple-point geostatistical models. In: O Leuangthong and C.V. Deutsch (eds.), *Geostatistics Banff*, vol 14, pp 235–244
- Santibañez † F**, Silva JF, Ortiz JM (2019) Sampling strategies for uncertainty reduction in categorical random fields: Formulation, mathematical analysis and application to multiple-point simulations. *Mathematical Geosciences* 51:579–624
- Santibañez-Leal † F**, Silva JF, Ortiz JM (2019) Improving short term planning and grade control with an entropy-based adaptive sampling strategy. *Natural Resources Research online first*:1–35
- Vapnik V (1999) *The Nature of Statistical Learning Theory*. Springer - Verlag, New York
- Vasát R, Heuvelink G, Borůvka L (2010) Sampling design optimization for multivariate soil mapping. *Geoderma* 155:147–153
- Verly G (2005) Grade control classification of ore and waste: a critical review of estimation and simulation based procedures. *Mathematical Geology* 37:417–476
- Vershynin R (2012) *Compressed Sensing: Theory and Applications*, 1st edn, Cambridge Univ Press, chap 5. Introduction to the non-asymptotic analysis of random matrices, pp 210–268
- Wellmann JF (2013) Information theory for correlation analysis and estimation of uncertainty reduction in maps and model. *Entropy* 15:1464–1485
- Wellmann JF, Regenauer-Lieb K (2012) Uncertainties have a meaning: Information entropy as a quality measure for 3-d geological models. *Tectonophysics* 526(Supplement C):207 – 216, doi: <https://doi.org/10.1016/j.tecto.2011.05.001>, URL <http://www.sciencedirect.com/science/article/pii/S0040195111001788>, modelling in Geosciences
- Wellmann JF, Horowitz FG, Schill E, Regenauer-Lieb K (2010) Towards incorporating uncertainty of structural data in 3d geological inversion. *Tectonophysics* 490(3):141 – 151, doi: <https://doi.org/10.1016/j.tecto.2010.04.022>, URL <http://www.sciencedirect.com/science/article/pii/S0040195110001691>
- Wellmer FW (1998) *Statistical Evaluations in Exploration for Mineral Deposits*. Springer-Verlag
- Wu J, Boucher A, Zhang T (2008) A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. *Comput Geosci* 34(12):1863–1876

Xu C, Hu C, Liu X, Wang S (2017) Information entropy in predicting location of observation points for long tunnel. *Entropy* 19(7), doi: 10.3390/e19070332, URL <http://www.mdpi.com/1099-4300/19/7/332>

Yeung RW (2002) *A First Course in Information Theory*. Springer

Zhang C, Li W (2008) A comparative study of nonlinear markov chain models for conditional simulation of multinomial classes from regular samples. *Stochastic Environmental Research and Risk Assessment* 22(2):217–230, doi: 10.1007/s00477-007-0109-2, URL <http://dx.doi.org/10.1007/s00477-007-0109-2>

Zidek J, Sun W, Le D (2000) Designing and integrating composite networks for monitoring multivariate gaussian pollution fields. *Appl Stat* 49:63–79

Nomenclature

\mathcal{A}	Alphabet
C_k	Resolvability Capacity
C_v	Covariance Matrix
$H(X)$	Shannon's entropy of X
f	Set of k measurements
facies	Rock structures recognized by its composition or fossil content and mapped by these characteristics
f^c	Complement of f
\mathbf{F}_k	Set of all f of size k
$f(l)$	l -th measurement in f
\mathbb{P}_X	Joint Probability of X
k	Number of measurements
$\mathbb{P}_{X_{ij}}$	Marginal Probability of X_{ij}
N	1D random array size
$[N] \times [N]$	Available positions set for a square 2-D image
Rm	Metallurgical Recovery
X	Regionalized Variable (R.V.)
X_f	X subset within f
$X_{i,j}$	R. V. in (i, j) position

Special Terms

ADA	ADaptive sampling approach.
AdSEMES	Adaptive Sequential Empirical Maximum Entropy Sampling.
AEP	Asymptotic Equipartition Property.
AMIS	Adaptive Maximum Information Sampling.
bGA	binary Genetic Algorithm.
BP	Basis Pursuit.
BPDN	Basis Pursuit DeNoising.
cGA	continuous or real-valued GA.
COg	<i>Cut-Off</i> grade.
CS	Compressive Sensing.
DCT	Discrete Cosine Transform.
FDG	Finite Difference Gradient.
GA	Genetic Algorithm.
GP	Gaussian Process.
GSLIB	Geostatistical Software Library.
HD	Hard Data (i.e. Samples or measurements).
i.d.	Identically Distributed.
i.i.d.	Independent Identically Distributed.
IT	Information Theory (or Information Theoretic).
LHS	Left Hand Side of an equation.
MC	Multiple Channel.
MES	Maximum Entropy Sampling.
MI	Mutual Information.
MMSE	Minimum Mean Square Error estimator.
MP	Matching Pursuit.
MPS	Multiple point simulations.

MRF	Markov Random Field.
NCS	Noisy Compressive Sensing.
nipd	No Inter-Pixel Dependency.
NP	Non deterministic Polynomial time.
OMP	Orthogonal Matching Pursuit.
OSP	Optimal Sensor Placement.
OWP	Optimal Well Placement.
pdf	Probability Density Function.
pmf	Probability Mass Function.
PSO	Particle Swarm Optimization.
RAMIS	Regularized Adaptive Maximum Information Sampling.
RHS	Right Hand Side of an equation.
RIP	Restricted Isometry Property.
RV	Regionalized Variable.
SC	Single Channel.
SD	Soft Data (i.e. weak samples and simulated data).
SGEMS	Stanford GEostatistical Modeling Software.
SMIS	Sequential (non-adaptive) Maximum Information Sampling.
SNESIM	Single Normal Equation SIMulation.
SNR	Signal to Noise Ratio.
SPSA	Simultaneous Perturbation Stochastic Approximation.
SSIM	Structural Similarity Index Measure.
STR	STRuctured sampling approach.
TI	Training Image.
TV	Total Variations.
VFSA	Very Fast Simulated Annealing.