Chapter 2

Research Proposal

The *Geosciences* global outcome correspond to the characterization of an unknown complex reservoir field. However, on account of the high number of structures, facies and properties in the subsurface, *geologists* study the field through the description of its individual components.

Thus, we focus on the recovery of a single subsurface property described by a *2-D* regularized variable sampled in the spatial domain. The characterization of these kinds of structures is relevant for mineral prospecting, mining planning, and production stages where the number of available observations is severely restricted by technical and economic factors. Based on the above, the integration of adaptive sensing schemes and sparse promoting approaches is a promising solution to improve classical geostatistical analysis.

2.1 Hypotheses

First, we summarize the main questions related with the channelized signal recovery in low acquisition regimes.

2.1.1 Questions

- What is the minimal observations amount that allows us to achieve an appropriate field characterization?
  - For global characterization, *Geosciences* two-points statistics approaches are useful for high sampling regimes.
  - For full characterization (i.e. description of local details), both sensing design tools and recovery methods based on *sparsity promoting* solvers permit a near perfect reconstruction from low acquisition regimes.

- Given K available measures, what is the *best* location for each one?
  - Sparse promoting approaches provide near perfect reconstructions of sparse signals under specific random sensing schemes.
  - Sampling design theory state that these locations are defined by field complexity and spatial statistics.

2.1.2 Hypotheses

The main hypotheses for this proposal are the following:

- The incorporation of prior information in the design of sampling schemes improves the performance of classical geosciences approaches at low acquisition regimes.
- A specific prior information can be used from different ways involving the achievement of better outcomes by its simultaneous incorporation in sensing design and sparse recovery tasks.
- Prior information based on multipoint simulations (*MPS*) conditioned to hard data (sampled data) allows us to estimate the spatial correlation of binary regionalized variables, and the implementation of adaptive sampling schemes.
- The estimation of the spatial correlation and the incorporation of additional constrains on spatial dependence allows us to approximate covariance keeping the problem computationally tractable.
- Training images provide statistical estimation (i.e. empirical *pdf*s) by pattern occurrences analysis under the assumption of stationarity. At non-stationary scenarios, *MPS* works as an optional way to access to the *pdf*s estimation.
- Adaptive Compressive Sensing (*ACS*) allows us to integrate sensing design and sparse recovery tools together to improve reconstruction algorithms.

2.1.3 Main Objective

The main objective for this research proposal is the development of methods to assist the reconstruction of images describing *2-D* binary regionalized variables by the use of *sensing design* tools, *sparse promoting* techniques and side information from spatial correlation analysis. First, we take advantage of information theoretical tools for the formulation and implementation of adaptive sensing schemes. We focus on regionalized variables with several spatial dependence assumptions by taking advantage of spatial structure and other side knowledge of interest media. Second, given several sensing schemes, we evaluate the sparse promoting techniques addressing practical limitations related with the problem complexity. In addition, a comprehensive study of the integration of both adaptive sampling and recovery methods is proposed by adaptive compressive sensing theory.

2.1.4 Specific Objectives

The specific objectives of this research proposal are:

- Formalize a theoretical framework for *2-D* regionalized variables imposing several spatial dependence constrains in the image model.
- Develop an adaptive sensing design framework using *joint entropy* and *mutual information* to measure uncertainty and spatial structure.
- Study a family of Markov random field models to describe spatial correlation on finite alphabet regionalized random variables.
- Study empirical statistic sources (*MPS* and training images) to estimate the spatial dependence.
- Incorporate a noisy sparse promoting recovery framework by the implementation of *NCS* principles oriented to evaluate its effect on the sampling schemes provided by *OWP*.
- Incorporate virtual noisy measurements in the proposed sparse recovery framework using *MPS* realizations and its empirical covariance as a virtual observations source.
- Integrate the concepts of sensing design and sparse representations on an adaptive compressive sensing approach for *2-D* binary permeability channels reconstruction.

2.2 Methodology

This work will be mainly based on the following steps: data base generation and consolidation applied to channelized structures in geosciences [8, 18, 24, 36], spatial correlation estimation for *NCS* [34, 33], formulation of sampling design problem [23, 13, 14, 11], incorporation of noisy sparse promoting solvers [3, 35, 37] and integration of the sensing and recovery proposed frameworks.

At next sections we describe the proposed methodology required to achieve the objectives posted on sections 2.1.3 and 2.1.4.

2.3 Data Base Definition

2.3.1 Data Base Generation

Here, we identify the models and phenomena of interest to be studied in this proposal. Three kinds of binary channels are proposed as shown in fig. 2.1. These channelized structures are obtained by unconstrained simulations using the geostatistical software *SGeMS* by the *SNESim* algorithm [18, 36].

Figure 2.1: Types of channels proposed for experimental analysis on this thesis.
From left to right: Single channel example (SC1), Multi channel 1 example (MC1), Multi channel 2 example (MC2)

2.3.2 Resizing Data Base

To build the database of permeability channels we consider *2-D* images of size $200x200$. It is an important dimension of the problem setting because the proposed approaches involve optimization problems where computational restrictions need to be imposed.

2.4 Estimation of Spatial Correlations

As the subsurface models of interest (i.e. channelized structures) has spatial correlations, regionalized random fields provide an appropriate framework to capture these spatial dependencies of our models [22]. In this context, we propose to study the spatial distribution as a function of *information theoretic* measures. These *information theoretic* measures will be adapted as a formal criteria to support decision making for *OWP* problem and to describe the remaining uncertainty after the measurements. This work relies on recent contributions [12] concerning the use of information measures for the analysis of geological random functions in terms of uncertainty and spatial dependencies.

We propose the use of statistical analysis from *MPS* realizations as an estimation of field uncertainty under spatial dependence assumption. As exposed in fig. 2.2, we use initial samples from the realization field (a selected realization image from the database) and one previously obtained training image to induce some coherent patterns structure in the simulations[1].

[1] the importance and selection of training images was described in section 1.4.3

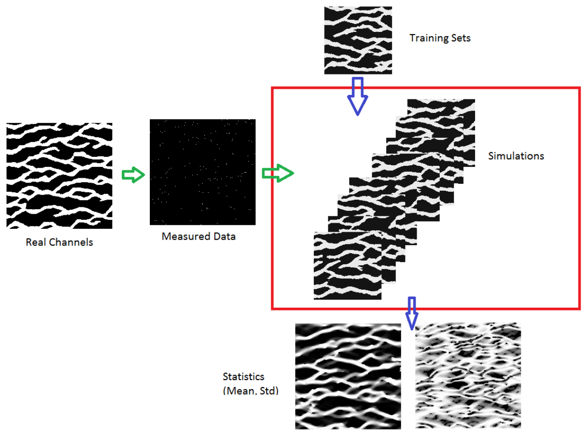

Figure 2.2: Scheme of the estimation of field statistics by the use of some hard data measurements and conditioned simulations using *MPS*.

Then, given a set of conditioned simulations we estimate the mean and variance of the individual random variables on the channelized field from it set.

2.5 Formulation of the Optimal Well Placement Problem

As discussed above in section 1.5, and let K be the number of possible measurements, then the main question that we want to solve is: where do you place these measurements?. In a preliminary approach, we state our problem through stochastic theory and *information theoretic* tools.

2.5.1 Joint Entropy Optimization for OWP

First, we formalize the problem considering the media of interest as a simplified *2-D* finite regionalized variable by the use of a finite alphabet image representing the subsurface channels.

The object to be characterized is a random image (or random field) denoting the subsurface distribution by a collection of finite alphabet random variables $X = \{X_i : i \in [N]\}$ where $[N] = \{1, \ldots, N\}$. For every position i in the array, X_i is a random variable with values in a finite alphabet $A^{|X_i|}$.

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

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

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

Adopting the concept of entropy as a measure of uncertainty of a random variable [38], we propose an algorithm that finds the placement rule f through optimal reduction of *a posteriori* entropy. The criteria used in Eq. (2.1) states that the measurement of the most uncertainty set of K positions will provide an optimal global reduction of the uncertainty for the media of interest (from the point of view of *information theory*).

$$X_f^* = \arg\max_{X_f \subset X} H(X_f) \tag{2.1}$$

More precisely, we try to characterize the conditional posterior entropy that in this context can be expressed as the joint entropy of the entire process minus the joint entropy of the variables measured by f, as shown in Eq.(2.2).

$$H(X^f | X_f) = H(X) - H(X_f) \tag{2.2}$$

Note that objective function in Eq. (2.1) for the search algorithm can use either the left $(H(X^f|X_f))$ or right $(H(X) - H(X_f))$ side of Eq.(2.2).

2.5.2 Formulation and Algorithmic Solution for the OWP

Over the collection of decision rules $\mathbf{F}_K$, we propose to chose the rule that minimizes, in average, the remaining uncertainty after taking the measurements, or the uncertainty of the remaining variables conditioned by the measured variables. More precisely, given a sensor placement decision rule $f \in \mathbf{F}_K$ let us denote by:

$$X_f \equiv (X_{f(1)}, X_{f(2)}, \dots, X_{f(K)}) \tag{2.3}$$

the measured random vector and by,

$$X^f \equiv (X_{\text{i}} : \text{i} \in [N] \setminus f) \tag{2.4}$$

the non-measured random vector. Note that f subset denotes all the spatial points measured by the *OWP* rule and $f^c \equiv [N] \setminus f$ its complement. Then considering a specific measure $X_f = (X_{f(1)}, \ldots, X_{f(K)}) = x_f \in \mathcal{A}^K$, the remaining uncertainty can be quantify by the *Shannon entropy* [38] of X^f given X_f, *i.e.* $H(X^f|X_f = x_f)$. Note that $H(X^f|X_f = x_f)$ represents the uncertainty conditioning to the specific measured values $(x_{f(1)}, \ldots, x_{f(K)})$. In practice, we do not have access to this measurements while making a decision on $\mathbf{F}_K$. Consequently, our objective function should consider the posterior uncertainty in average with respect to the statistics of X_f. In other words, we consider the *Shannon conditional entropy* [38] of X^f given X_f, *i.e.*:

$$H(X^f|X_f) = -\sum_{x_f \in A^K} P_{X_f}(x_f) H(X^f|X_f = x_f) \tag{2.5}$$

as the objective function. Then the *OWP* of K-measurements reduces to:

$$f_K^* \equiv \arg\min_{f \in \mathbf{F}_K} H(X^f|X_f) \tag{2.6}$$

which is the solution that minimizes the posterior uncertainty.

2.5.3 Estimation of the Stochastic Field Model

The proposed *OWP* formulation is based on the knowledge of the statistics for the random regionalized variables. In practice, this theoretical model is not available and we need to find the way to estimate this model from empirical data. Sources of empirical statistics are reference images provided by experts, historical information or imposed models. In this work the use of training images and geostatistical simulation tools is proposed as a source for estimating the joint and conditional distributions of X.

2.5.4 Scheme of Sampling Design and Experimental Validation

A basic scheme of the inference process including preliminary blind sampling and adaptive *OWP* is illustrated in fig. 2.3.

The implementation of the different blocks of the proposed system at fig. 2.3 will lead to several practical variants to take advantage of available external information sources. In order to evaluate the *OWP* solutions a performance comparison with random and structured sampling schemes will be performed.

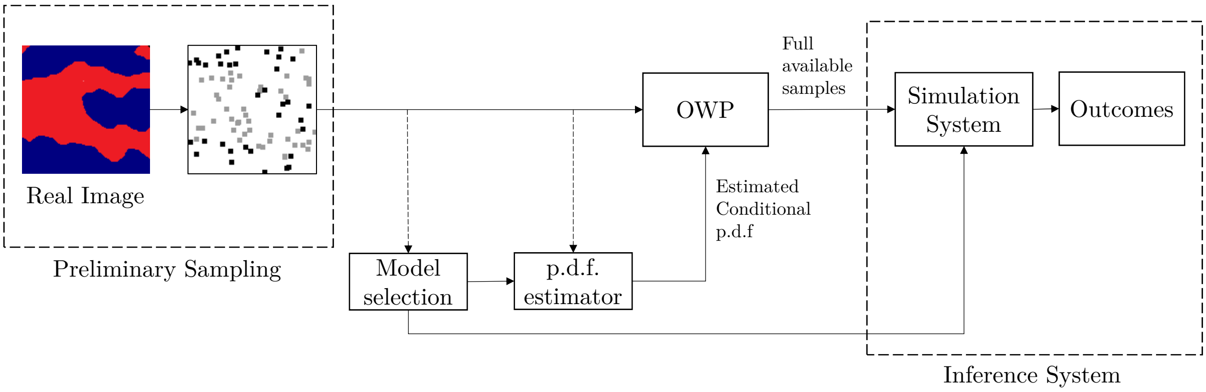

Figure 2.3: General scheme used

2.6 Formulation and Implementation of Noisy Sparse Promoting Solvers

Given sampling schemes provided by *OWP*, we propose its comparison with some sparse promoting oriented schemes by the modification of classical $L1$ minimizers solver (available in the $L1$ magic and *CVX* software). Our contribution focus on formulating the signal recovery process as a generalized sampling problem. The idea is 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, we apply principles from Noisy compressive Sensing *NCS* to the reconstruction of binary channels of permeability. For the estimation of the model, we propose the use of *MPS* as a source of statistical data. In particular, it allows the estimation of variance and covariance of the target regionalized variables.

2.6.1 *NCS* and *Whitening* process

We incorporate the noise associated to measurements 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 $m << N$), the target is to reconstruct the *real* channel. While traditional *CS* approaches deal with time-invariant sparse signals without error in measurements, our 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, we implemented and validated a preliminary framework of *NCS* oriented to improve *MPS* performance. Standard *CS* implementation for channelized binary structures was proposed and implemented for others members of the *IDS* Lab [2]. The reader is referred to [39] for more details on this approximation. Classical conditions and theory of *NCS* was described in sections 1.6.2 and 3.4.2.

Here, we extend previous works in order to consider noisy measures and provided a framework to noise characterization. In order to apply the theory of *NCS* we requires a signal model with white noise, then a whitening process is required to use the proposed methods. Thus, we consider the next *sensing model* for our problem:

$$I = \Phi \cdot Z \tag{2.7}$$

$$X = I(:) \tag{2.8}$$

$$Y = A \cdot X + \xi \tag{2.9}$$

with

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

At this point the eq. (2.9) only differs from [39] approach in the incorporation of noise component. Here, the noise ξ would be a spatially correlated noise, then we required a *whitening* pre-processing.

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

$$Y - \xi_{mean} = A \cdot X + \xi - \xi_{mean} \tag{2.10}$$

Defining zero mean variables and rewriting we obtain the expression in eq. (2.11):

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

$$Y_0 = A \cdot X + \xi_0 \tag{2.11}$$

From eq. (2.11) 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 we achieved the next formulation:

$$C_v^{\frac{-1}{2}} \cdot Y_0 = C_v^{\frac{-1}{2}} \cdot A \cdot X + C_v^{\frac{-1}{2}} \cdot \xi_0 \tag{2.12}$$

$$C_v^{\frac{-1}{2}} \cdot Y_0 = C_v^{\frac{-1}{2}} \cdot A \cdot X + \eta \tag{2.13}$$

$$\widehat{Y} = C_v^{\frac{-1}{2}} \cdot A \cdot X + \eta \tag{2.14}$$

Replacing the vectorization process we finally obtain the next model:

$$\widehat{Y} = C_v^{\frac{-1}{2}} \cdot A \cdot VEC(\Phi \cdot Z) + \eta \tag{2.15}$$

The eq. (2.15) 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*.

2.7 Performance Metrics

The performance metrics consider field reconstruction quality analysis and variability of conditioned simulations.

For reconstruction, comparisons between the original image and the reconstructions will be performed. Classical indicators assess the quality of a *2-D* image, X, by comparing it with a reference *2-D* image, Xr, [40] with both images of size M_u by M_v.

The signal to noise ratio *SNR* expressed in decibels dB:

$$\mathbf{SNR} = 10 \cdot log_{10} \left(\frac{\sum_{u=1}^{M_u} \sum_{v=1}^{M_v} |Xr_{u,v}|^2}{\sum_{u=1}^{M_u} \sum_{v=1}^{M_v} |Xr_{u,v} - X_{u,v}|^2} \right) \tag{2.16}$$

The peak signal to noise ratio *PSNR* expressed in decibels dB:

$$\mathbf{PSNR} = 10 \cdot log_{10}\left(\frac{max(Xr_{u,v})^2}{\frac{1}{M_u \cdot M_v} \cdot \sum_{u=1}^{M_u}\sum_{v=1}^{M_v}|Xr_{u,v} - X_{u,v}|^2}\right) \tag{2.17}$$

The root mean square error $RMSE$:

$$\mathbf{RMSE} = \sqrt{\frac{1}{M_u \cdot M_v} \cdot \sum_{u=1}^{M_u}\sum_{v=1}^{M_v}|Xr_{u,v} - X_{u,v}|^2} \tag{2.18}$$

The mean absolute error MAE:

$$\mathbf{MAE} = \frac{1}{M_u \cdot M_v} \cdot \sum_{u=1}^{M_u}\sum_{v=1}^{M_v}|Xr_{u,v} - X_{u,v}| \tag{2.19}$$

On the other hand, the structural similarity index $SSIM$ tries to estimate the similarity between two images by emulating the human perception [41]. A simplified version of this indicator is given by:

$$\mathbf{SSIM}(X, Xr) = \frac{(2\mu_X\mu_{Xr} + c_1)(2\sigma_{X,Xr} + c_2)}{(\mu_X{}^2 + \mu_{Xr}^2 + c_1)(\sigma_X^2 + \sigma_{Xr}^2 + c_2)} \tag{2.20}$$

Here, μ_X (μ_{Xr}) and σ_X (σ_{Xr}) denote the average and variance of X (Xr), respectively. $\sigma_{X,Xr}$ is the covariance of X and Xr. c_1 and c_2 are required to stabilize the division.

For our problems, outcomes are the sensing locations provided for each method. Then, we don't have any reconstruction in order to compare different approaches. Thus, an alternative quality analysis based on the variability of simulations conditioned by the obtained sensing locations is proposed. Then we can estimate the mean and standard deviations from simulations for each implemented approach.

Another analysis is to evaluate the global behavior of simulations in relation with the target field. We propose the SNR indicator between the mean of simulations and the reference image and the mean of the SNR of each simulation and the reference image. Finally we propose to use some reconstruction method from data measured at sensing locations like total variation inpainting.