Chapter 3

Preliminary Results

As part of preliminary work we evaluated inverse problems approaches applied to *Geostatistics subsurface channels* and *Imaging*. We organize the preliminary outcomes in the next sections as theoretical or formulation stages and synthetic experimental stages.

In addition, in the current year we are consolidating the developed approaches writing a manuscript related with *OWP* in *2-D* binary channels of permeability.

3.1 On Optimal Well Placement Formulation

There is an interesting interpretation of the eq. (2.6) (in section 2.5), as a problem of maximum information extraction. In this section we summarize this interpretation and study various suboptimal approaches. Furthermore, we propose a field complexity indicator based on *OWP* and show the behavior of this indicator in some typical distributions.

3.1.1 The Equivalent Maximum Information Decision

Adopting the well-known concept of mutual information in information theory [38], we can define the amount of information that provides a rule $f \in \mathbf{F}_K$ as the reduction of the uncertainty after taking the measurements. More precisely, we define the information of f to resolve X as:

$$I(f) \equiv H(X) - H(X^f|X_f) \tag{3.1}$$

where $H(X)$ is the Shannon entropy of the whole array X (or the *a priori* uncertainty) given by:

$$H(X) = -\sum_{x \in A^N} P_x(X = x) \cdot \log P_X(X = x). \tag{3.2}$$

$I(f)$ is a particular case of the well-known mutual information [38], which implies that $I(f) \geq 0$ and $I(f) \leq H(X)$ [1]. Then our *OWP* problem from eq. (2.6) can be posted, using eq. (3.1), as the problem of maximizing the information to resolve X with K-measurements:

$$f_K^* = \arg\max_{f \in \mathbf{F}_K} I(f). \tag{3.3}$$

To conclude the formulation, we derive a final interpretation of the decision problem in eq. (3.1) and eq. (3.3) using the chain rule of the entropy [38]. The chain rule tells us that for any $f \in \mathbf{F}_K$ we can decompose the joint entropy $H(X)$ as the sum of the marginal entropy of the sensed pixels $H(X_f)$ and the conditional entropy $H(X^f|X_f)$. Using that in eq. (3.1) and then in eq. 3.3 we have that:

$$f_K^* = \arg\max_{f \in \mathbf{F}_K} H(X_f), \tag{3.4}$$

which is the problem of choosing the K-positions that has the highest *a-priori* joint entropy. Then our original problem of minimizing the posterior Shannon entropy after the selection of the sensing positions in eq. (2.6) is equivalent to the problem of finding the K positions that maximizes the information to resolve X in eq. 3.3, which is also equivalent to find the subset of K-measurements that maximizes the *a-priori* uncertainty before taking the measurements in eq. (3.4).

3.1.2 Resolvability Capacity of $\{X_{\mathrm{i}}\}$

Before moving to next sections, it is interesting to analyze an indicator of the complexity of the field in terms of the capacity to resolve its uncertainty with K oriented measurements. We define the resolvability capacity of $\{X_{\mathrm{i}}\}$ with K-measurements as:

$$\mathcal{C}_K \equiv \frac{I(f_K^*)}{H(X)} \in [0, 1]. \tag{3.5}$$

for all $K \in [N]$. This is the ratio between the information gain of the best sensor-placement rule in eq. (3.3) 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 we can show the following properties:

PROPOSITION 1 For any arbitrary random field $\{X_{\mathrm{i}}\}$:

1. $\mathcal{C}_{K+1} \geq \mathcal{C}_K$, $\forall K \in \{1, \ldots, N-1\}$.

2. $\mathcal{C}_N = 1$ and we can define $\mathcal{C}_0 = 0$.

[1] If $I(f) = H(X)$ implies that $H(X^f|X_f) = 0$, which is equivalent to say that X^f is a deterministic function of X_f [38] and, consequently, the sensing rule f perfectly resolves X with no uncertainty.

Hence, $\{\mathcal{C}_K : K\,\{1,\ldots,N\}\}$ 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 [2] measurements.

3.1.3 Iterative sub-optimal solution for *OWP*

The *OWP* problem, as presented in Eq.(2.6) is combinatorial and, consequently, impractical for relatively large random images [30]. We propose 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 algorithm (to something polynomial in the size of the problem) and, therefore, implementable [42, 22, 17].

Let us begin with $K = 1$. In this context the *OWP* problem reduces to find one position in the array solution of:

$$(\mathrm{i}_1^*) = \underset{(\mathrm{i})\in[N]}{\arg\max}\, H(X_{\mathrm{i}}) \tag{3.6}$$

by Eq.(2.2).

The idea of the one-step ahead approach implies to fix i_1^* and then find the next position in $[N] \setminus \{\mathrm{i}_1^*\}$ solution of:

$$(\mathrm{i}_2^*) = \underset{\substack{(\mathrm{i})\in[N]\\(\mathrm{i})\neq(\mathrm{i}_1^*)}}{\arg\max}\, H(X_{\mathrm{i}}|X_{\mathrm{i}_1^*}) \tag{3.7}$$

Then the problem in Eq.(3.7) finds the point that maximizes the *a priori* uncertainty (conditioning to previous measured data, in this case (i_1^*)), which is the simplest principle to implement.

Hence, iterating this inductive (sensing) rule using the chain-rule of the entropy [38], we have that the k-measurement (after taking $(\mathrm{i}_1^*), (\mathrm{i}_2^*), \ldots, (\mathrm{i}_{k-1}^*)$ is the solution of:

$$(\mathrm{i}_k^*) = \underset{\substack{(\mathrm{i})\in[N]\\(\mathrm{i})\neq(\mathrm{i}_p^*)\\p=1,\ldots,k-1}}{\arg\max}\, H(X_{\mathrm{i}}|X_{\mathrm{i}_1^*},\ldots,X_{\mathrm{i}_{k-1}^*}) \tag{3.8}$$

Therefore with this sequence of ordered data points $\{(\mathrm{i}_k^*) : k = 1,\ldots,K\}$, for every $K \in [N]$ we can construct the iterative sensor-placement rule $\tilde{f}_K^* \in \mathbf{F}_K$ by $\tilde{f}_{K(1)}^* = (\mathrm{i}_1^*), \tilde{f}_{K(2)}^* = (\mathrm{i}_2^*), \ldots, \tilde{f}_K^*(k) = (\mathrm{i}_k^*)$.

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 $X_{\mathrm{i}_1^*}$), which is the simplest principle to implement.

[2] in the information theoretic meaning.

Concerning the information of this iterative scheme, we can state that:

PROPOSITION 2 The information of $\tilde{f}_K^*$ to resolve the field X (using the definition in eq. 3.1) is given by:

$$\begin{aligned} I(\tilde{f}_K^*) &= H(X) - H(X^{\tilde{f}_K^*}|X_{\tilde{f}_K^*}) \\ &= H(X_{\mathrm{i}_1^*}) + H(X_{\mathrm{i}_2^*}|X_{\mathrm{i}_1^*}) + \cdots + H(X_{\mathrm{i}_K^*}|X_{\mathrm{i}_1^*}, \ldots, X_{\mathrm{i}_{k-1}^*}). \\ &= H(X_{\mathrm{i}_1^*}, \ldots, X_{\mathrm{i}_{k-1}^*}, X_{\mathrm{i}_K^*}), \end{aligned} \tag{3.9}$$

and, consequently, the information gain of this iterative scheme is *additive* in the sense that:

$$I(\tilde{f}_K^*) - I(\tilde{f}_{K-1}^*) = H(X_{\mathrm{i}_K^*}|X_{\mathrm{i}_1^*}, \ldots, X_{\mathrm{i}_{k-1}^*}) \geq 0. \tag{3.10}$$

Note that the information gain from $K-1$ to K measurements in eq. (3.10) derives directly from the solution of eq. (3.8).

3.1.4 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 $K \in \{1, \ldots, N\}$

$$I(\tilde{f}_K^*) \leq I(f_K^*). \tag{3.11}$$

Consequently, a concrete way to evaluate how much we loss in the reduction of the uncertainty between the combinatorial optimal scheme $\{f_K : K \geq 1\}$ and the practical iterative scheme $\left\{\tilde{f}_K^* : K \geq 1\right\}$, is given by the difference between $\mathcal{C}_K$ in eq. 3.5 and

$$\tilde{\mathcal{C}}_K \equiv \frac{I(\tilde{f}_K^*)}{H(X)} \in [0, 1]. \tag{3.12}$$

We conjecture that the information loss $(\mathcal{C}_K - \tilde{\mathcal{C}}_K)_{K=1,\ldots,N}$ 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., $P_X(X)_{(\mathrm{i})\in[N]}) = \Pi_{(\mathrm{i})\in[N]} P_{X_\mathrm{i}}(X_\mathrm{i})$, we have that:

$$\mathcal{C}_K = \tilde{\mathcal{C}}_K, \tag{3.13}$$

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

3.1.5 OWP and Resolvability Capacity Properties

In order to quantify the level of *knowledge* on the media given an optimal sampling $f_K^* = \{(\mathrm{i}_1), \ldots, (\mathrm{i}_K)\}$ we resume the next properties for $C_K = \dfrac{H(X_{f_K^*})}{H(X)}$:

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

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

$$H(X_f) = \sum_{\mathrm{i} \in f} H(X_\mathrm{i}) \overset{\text{(Only IID)}}{=} \mid f \mid \cdot H(X_{\mathrm{i}^*}) \tag{3.14}$$

and the resolvability capacity is reduced to:

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

In real scenarios the field under analysis incorporates some level of spatial interdependence. A case of interest for our current and future work is associated with Markov random fields *MRFs* where we can characterize the spatial dependence by Markov chains and to define spatial scope of interaction by the incorporation of *Cliqués*. In simulation of fields with spatial dependence *MRFs* provide an useful tool [43] as illustrated in figure 3.1 but we are focused in the use of *MRFs* to empirically estimate the spatial dependence in the field.

By the use of approximations of *Cliqués* we can describe joint probability models by considering a markovian property. For example, if we define $b_{u,v} = x_\mathrm{i}$ with $\mathrm{i} \in [N]$ and $(u, v) \in [M]x[M]$ as a mapping transform (from our original stochastic field space without specific spatial order to a sorted *2-D* space) and assuming a *Cliqué* with only vertical and horizontal dependence of unitary distance $(\mathcal{N}(b_{u,v}) = \{b_{u-1,v}, b_{u,v-1}, b_{u+1,v}, b_{u,v+1}\})$ we can reduce the joint probability of a specific spatial variable by only considering the interaction with the members of it *Cliqué* (only four variables in the neighborhood of this example). A realistic characterization of the *Cliqué* will provides better estimation for the joint entropy while a simplified version of the *Cliqué* will offer more computationally affordable implementations.

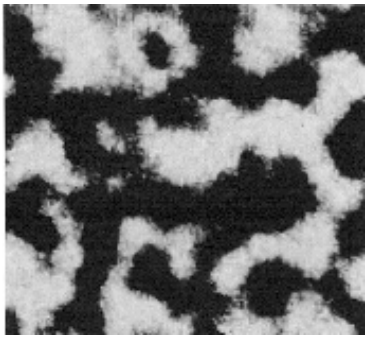

Figure 3.1: synthetic image by (MRF) modeling [43]

We want to explore in detail this kind of markovian models in the next stages of this thesis.

3.1.6 Recapitulation and interpretation of implemented *OWP* rules

An optimal (from *information theoretic* approach) sensor placement rule for select K measurements rely on the maximization of the prior entropy $H(X_f)$ with $f \in F$. By chain rule for entropy:

$$H(X_f) = H(X_{f(1)}) + H(X_{f(2)}|X_{f(1)}) + \ldots + H(X_{f(K)}|X_{f(K-1)}, \ldots, X_{f(1)},) \tag{3.15}$$

In preliminary formulation we considered the case of the initial K measurements without previous sampled positions in the field. As iterative OWP rules take measurements one by one or in groups we can rewrite the entropy isolating the positions previously measured. Thus, the expression of eq. (3.16) can be divided in two parts. Let j be an index that accounts for the position of the J variables that have been previously measured by a rule $\widetilde{f}$, before the current optimization process f of $L = |f|$ new sampling positions is applied. Then,

$$H(X_f) = \sum_{j=1}^{J} H(X_{\widetilde{f}(j)}|X_{\widetilde{f}(j-1)}, \ldots, X_{\widetilde{f}(1)}) + \sum_{l=1}^{L} H(X_{f(l)}|X_{f(l-1)}, \ldots, X_{f(1)}, X_{\widetilde{f}(J)}, X_{\widetilde{f}(J-1)}, \ldots, X_{\widetilde{f}(1)}) \tag{3.16}$$

The first sum is equal to zero, because it represent the uncertainty of a set of variables that are known already by the initial J measurements. In addition, in the second sum we can recognizes two kind of conditioning variables as exposed in the definition Eq. (3.17):

$$X_{S_l} = \{X_{f(l-1)}, X_{f(l-2)}, \ldots, X_f(1), X_{\widetilde{f}(J)}, X_{\widetilde{f}(J-1)}, \ldots, X_{\widetilde{f}(1)}\} \tag{3.17}$$

While variables indexed from rule $\widetilde{f}$ from $1, \ldots, J$ are previous sampling positions and the specific realization for each variable could be accessible by the measurement process, the variables indexed by the rule f from $1, \ldots, L$ are the target of the current OWP rule.

For simplicity we can use a sorted combination $\widehat{f}$ of the rules $\widetilde{f}$ and f with indexes from $1, \ldots, K$. Indexes from $1, \ldots, J$ account the positions obtained from the rule $\widetilde{f}$ and the indexes from $J+1, \ldots, K$ accounts the positions obtained from the rule f.

$$X_{S_k} = \{X_{\widehat{f}(1)}, \ldots, X_{\widehat{f}(J)}, X_{\widehat{f}(J+1)}, \ldots, X_{\widehat{f}(K)}\} \tag{3.18}$$

Now, by the entropy chain rule we have rewritten the initial method, as the sum of marginal entropy of each wanted sensor position and its conditionals. This representation keeps the problem from original formulation: As we are doing an optimization over a space of size $\binom{L}{[N]-J}$ the solution to this problem can be found by applying an algorithm of combinatorial complexity which is unfeasible or at least highly computationally intensive for fields with thousands or more variables. Instead, we propose to analyze each element separately in an iterative algorithm.

$$X_f(\mathrm{i})^* = \arg\max_{X_\mathrm{i} \in X} H(X_\mathrm{i}|X_{S_\mathrm{i}}) \tag{3.19}$$

$$H(X_\mathrm{i}|X_{S_\mathrm{i}}) = \sum_{x_{S_\mathrm{i}} \in A^{|S_\mathrm{i}|}} P(X_{S_\mathrm{i}}) H(X_\mathrm{i}|X_{S_\mathrm{i}} = x_{S_\mathrm{i}}) \tag{3.20}$$

where x_{S_i} denotes the set of values that X_{S_i} takes in a particular realization. There is still a problem with the estimation of joint distribution for the variables we are optimizing on. To solve this problem, we can assume that $P(X_{S_\mathrm{i}})$ follows a degenerated distribution after X_{S_i} was measured with the realization $x^0_{S_\mathrm{i}}$. This assumption allows us to rewrite eq. (3.21) to:

$$H(X_\mathrm{i}|X_{S_\mathrm{i}}) = H(X_\mathrm{i}|X_{S_\mathrm{i}} = x^0_{S_\mathrm{i}}) \tag{3.21}$$

This new equation reduce the number of degrees of freedom in the joint distribution. Therefore, we only need to estimate the probabilities of X_i in the alphabet A conditioned to the specific $x^0_{S_\mathrm{i}}$ realization. This conditional entropy can be easily estimated using pattern search tools or *MPS* techniques for the case of *2-D* channelized structures studied in this thesis.

3.2 On Regionalized variables with spatial dependence

At this point we worked either without any kind of spatial dependence for random variables or only with very simple assumptions over it dependence for regionalized random variables. Currently we have interest in to develop and complement a formalization and practical reduction for it kind of dependence.

3.2.1 *MRF* models and *Cliqué* structure estimation

When the spatial dependence is headed by a markovian property we can reduce the conditional probabilities by only considering conditionals on the *Cliqué* of the variable of interest:

$$P(X_{\mathrm{i}}|X^{\mathrm{i}}) \equiv P(X_{\mathrm{i}}|X_{CL(\mathrm{i})}) \tag{3.22}$$

Where $CL(\mathrm{i})$ correspond to the *Cliqué* associated to the random variable X_{i} at the position i.

3.2.2 *Cliqué* estimation and Multi-Point/Two-Point Mutual Information relation

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

$$P(X_{\mathrm{i}}|X^{\mathrm{i}}) = f(\{P(X_{\mathrm{i}}|X_k);\ \forall k \in [N] \setminus \mathrm{i}\}) \tag{3.23}$$

Using Bayes rule we can write the following expression:

$$\begin{aligned} P(X_{\mathrm{i}}|X^{\mathrm{i}}) &= \frac{P(X_{\mathrm{i}}, X^{\mathrm{i}})}{P(X^{\mathrm{i}})} \\ &= \frac{P(X)}{P(X^{\mathrm{i}})} \\ &= \frac{P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}})P(X_{\{\mathrm{i},k\}})}{P(X^{\{\mathrm{i},k\}}|X_k)P(X_k)} \\ &= \frac{P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}})}{P(X^{\{\mathrm{i},k\}}|X_k)}P(X_{\mathrm{i}}|X_k) \end{aligned} \tag{3.24}$$

Eq. (3.24) shows that $P(X_{\mathrm{i}}|X^{\mathrm{i}})$ can be rewritten in terms of $P(X_{\mathrm{i}}|X_k)$ for any k in the complement subset. We can use this property to derive a combination of $P(X_{\mathrm{i}}|X_k)$ that is equal to multi-point conditional probability.

$$P(X_{\mathrm{i}}|X^{\mathrm{i}})^{|I_{X^{\mathrm{i}}}|} = \prod_{k \in I_{X^{\mathrm{i}}}} \frac{P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}})}{P(X^{\{\mathrm{i},k\}}|X_k)} P(X_{\mathrm{i}}|X_k) \tag{3.25}$$

Where $I_{X^{\mathrm{i}}}$ represent the set of indexes of X^{i}. Using the $|I_{X^{\mathrm{i}}}|$th-root in both sides of eq. (3.25), we have:

$$\underbrace{P(X_{\mathrm{i}}|X^{\mathrm{i}})}_{\text{multi-point prob.}} = \prod_{k \in I_{X^{\mathrm{i}}}} \underbrace{\left(\frac{P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}})}{P(X^{\{\mathrm{i},k\}}|X_k)} \right)^{\frac{1}{|I_{X^{\mathrm{i}}}|}}}_{\alpha_{\{\mathrm{i},k\}}} \underbrace{P(X_{\mathrm{i}}|X_k)}_{\text{two-point prob.}}^{\frac{1}{|I_{X^{\mathrm{i}}}|}} \tag{3.26}$$

That is what we wanted to get, as it was stated in eq. (3.23).

3.2.3 Multi-point and Two-Point Mutual Information relation

Applying the definition for mutual information and Eq. (3.26) we can write the following relations:

$$
\begin{aligned}
& I(X_{\mathrm{i}}; X^{\mathrm{i}}) \\
&= H(X_{\mathrm{i}}) - H(X_{\mathrm{i}}|X^{\mathrm{i}}) \\
&= H(X_{\mathrm{i}}) + \sum_{\{X_{\mathrm{i}},X^{\mathrm{i}}\}\in A^{|X|}} P(X_{\mathrm{i}}, X^{\mathrm{i}}) \log P(X_{\mathrm{i}}|X^{\mathrm{i}}) \\
&= H(X_{\mathrm{i}}) \\
&\quad + \sum_{\{X_{\mathrm{i}},X^{\mathrm{i}}\}\in A^{|X|}} P(X_{\mathrm{i}}, X^{\mathrm{i}}) \log \left(\prod_{k\in I_{X^{\mathrm{i}}}} \frac{P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}})}{P(X^{\{\mathrm{i},k\}}|X_k)} P(X_{\mathrm{i}}|X_k) \right)^{\frac{1}{|I_{X^{\mathrm{i}}}|}} \\
&= H(X_{\mathrm{i}}) \\
&\quad + \frac{1}{|I_{X^{\mathrm{i}}}|} \sum_{k\in I_{X^{\mathrm{i}}}} \sum_{\{X_{\mathrm{i}},X^{\mathrm{i}}\}\in A^{|X|}} P(X_{\mathrm{i}}, X^{\mathrm{i}}) \log \left(\frac{P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}})}{P(X^{\{\mathrm{i},k\}}|X_k)} P(X_{\mathrm{i}}|X_k) \right) \\
&= H(X_{\mathrm{i}}) \\
&\quad + \frac{1}{|I_{X^{\mathrm{i}}}|} \sum_{k\in I_{X^{\mathrm{i}}}} \sum_{\{X_{\mathrm{i}},X^{\mathrm{i}}\}\in A^{|X|}} P(X_{\mathrm{i}}, X^{\mathrm{i}}) \big(\log P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}}) \\
&\qquad\qquad + \log P(X_{\mathrm{i}}|X_k) \\
&\qquad\qquad - \log P(X^{\{\mathrm{i},k\}}|X_k) \big) \\
&= H(X_{\mathrm{i}}) + \frac{1}{|I_{X^{\mathrm{i}}}|} \sum_{k\in I_{X^{\mathrm{i}}}} \Bigg\{ \sum_{\{X_{\mathrm{i}},X^{\mathrm{i}}\}\in A^{|X|}} P(X_{\mathrm{i}}, X^{\mathrm{i}}) \log P(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}}) \\
&\qquad\qquad + \sum_{\{X_{\mathrm{i}},X^{\mathrm{i}}\}\in A^{|X|}} P(X_{\mathrm{i}}, X^{\mathrm{i}}) \log P(X_{\mathrm{i}}|X_k) \\
&\qquad\qquad - \sum_{\{X_{\mathrm{i}},X^{\mathrm{i}}\}\in A^{|X|}} P(X_{\mathrm{i}}, X^{\mathrm{i}}) \log P(X^{\{\mathrm{i},k\}}|X_k) \Bigg\} \\
&= H(X_{\mathrm{i}}) - \frac{1}{|X^{\mathrm{i}}|} \sum_{k\in I_{X^{\mathrm{i}}}} \big(H(X_{\mathrm{i}}|X_k) + H(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}}) - H(X^{\{\mathrm{i},k\}}|X_k) \big) \qquad (3.27)
\end{aligned}
$$

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

$$
\begin{aligned}
& H(X_{\mathrm{i}}|X_k + H(X^{\{\mathrm{i},k\}}|X_{\{\mathrm{i},k\}}) - H(X^{\{\mathrm{i},k\}}|X_k) \\
&\quad = H(X_{\mathrm{i}}) - I(X_{\mathrm{i}}; X_k) + H(X^{\{\mathrm{i},k\}}) - I(X_{\{\mathrm{i},k\}}; X^{\{\mathrm{i},k\}}) - H(X^{\{\mathrm{i},k\}}) + I(X_k; X^{\{\mathrm{i},k\}}) \\
&\qquad\qquad = H(X_{\mathrm{i}}) - I(X_{\mathrm{i}}; X_k) - I(X_{\{\mathrm{i},k\}}; X^{\{\mathrm{i},k\}}) + I(X_k; X^{\{\mathrm{i},k\}}) \qquad (3.28)
\end{aligned}
$$

Then, we can sort the expression of Eq. (3.27) in three kinds of components:

$$
\begin{aligned}
I(X_{\mathrm{i}}; X^{\mathrm{i}}) =& H(X_{\mathrm{i}}) \\
& - \frac{1}{|X^{\mathrm{i}}|} \sum_{k \in I_{X^{\mathrm{i}}}} H(X_{\mathrm{i}}) - I(X_{\mathrm{i}}; X_k) - I(X_{\{\mathrm{i},k\}}; X^{\{\mathrm{i},k\}}) + I(X_k; X^{\{\mathrm{i},k\}}) \\
=& \underbrace{\left[\frac{1}{|X^{\mathrm{i}}|} \sum_{k \in I_{X^{\mathrm{i}}}} I(X_{\mathrm{i}}; X_k) \right]}_{\text{one and two-point statistics}} \\
& + \underbrace{\left[\frac{1}{|X^{\mathrm{i}}|} \sum_{k \in I_{X^{\mathrm{i}}}} I(X_{\{\mathrm{i},k\}}; X^{\{\mathrm{i},k\}}) \right]}_{\text{higher order statistics}} - \underbrace{\left[\frac{1}{|X^{\mathrm{i}}|} \sum_{k \in I_{X^{\mathrm{i}}}} I(X_k; X^{\{\mathrm{i},k\}}) \right]}_{\text{redistributable term}}
\end{aligned} \tag{3.29}
$$

We can extend multi-point conditional probability for a lower order subset $X^{\{k,J\}}$where $J \subset [N]$ and $k \notin J$.

$$
P(X_k | X^{\{k,J\}}) = \left(\prod_{l \in I_{X^{\{k,J\}}}} \frac{P(X^{\{k,l,J\}} | X_{\{k,l\}})}{P(X^{\{k,l,J\}} | X_{\{l\}})} P(X_k | X_{\{l\}}) \right)^{\frac{1}{|I_{X^{\{k,J\}}}|}} \tag{3.30}
$$

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

$$
\begin{aligned}
I(X_k; X^{\{k,J\}}) = & \left[\frac{1}{|X^{\{k,J\}}|} \sum_{l \in I_{X^{\{k,J\}}}} I(X_k; X_{\{l\}}) \right] \\
& + \left[\frac{1}{|X^{\{k,J\}}|} \sum_{l \in I_{X^{\{k,J\}}}} I(X_{\{l,k\}}; X^{\{l,k,J\}}) \right] \\
& - \left[\frac{1}{|X^{\{k,J\}}|} \sum_{l \in I_{X^{\{k,J\}}}} I(X_{\{l\}}; X^{\{l,k,J\}}) \right]
\end{aligned} \tag{3.31}
$$

Using eq. (3.31) in eq. (3.29), recursively, we obtain:

$$
\begin{aligned}
I_{|X|}(X_{\mathrm{i}};X^{\mathrm{i}}) = &\left[2H(X_{\mathrm{i}}) + \mathcal{H}_{|X|-1}\sum_{\substack{k\in I_X\\ k\neq \mathrm{i}}} H(X_k)\right] \\
&- \left[\frac{1}{|X|-1}\sum_{k\in I_{X^{\mathrm{i}}}} I(X_{\mathrm{i}};X_k)\right. \\
&\left. + \left(\frac{N}{N-1} - \frac{1}{(N-1)!}\right)\left(\sum_{k\in I_{X^{\mathrm{i}}}}\sum_{j\in I_{X^{\{\mathrm{i},k\}}}} I(X_k;X_j)\right)\right] \\
&+ \left[\sum_{m=1}^{|X|-1}\frac{1}{(m-1)!}\sum_{\substack{s_1\in I_X\\ s_1\neq s_0}}\cdots\sum_{\substack{s_m\in I_X\\ s_m\neq s_0,\cdots,s_{m-1}}} I(X_{s_m,s_{m-1}};X^{s_0,\ldots,s_m})\right]
\end{aligned}
\tag{3.32}
$$

3.3 On Optimal Well Placement Performance

Based on our preliminary theoretical formalization for the near optimal measurements location, in the characterization of subsurface channelized fields of permeability, we implemented an iterative *OWP* solution of eq. (3.19) and provided a comparison between *OWP* and structured/randomized sampling schemes.

3.3.1 Implementation of *OWP* to *2-D* binary channels

Our scheme consider the stages exposed in the fig. 3.2

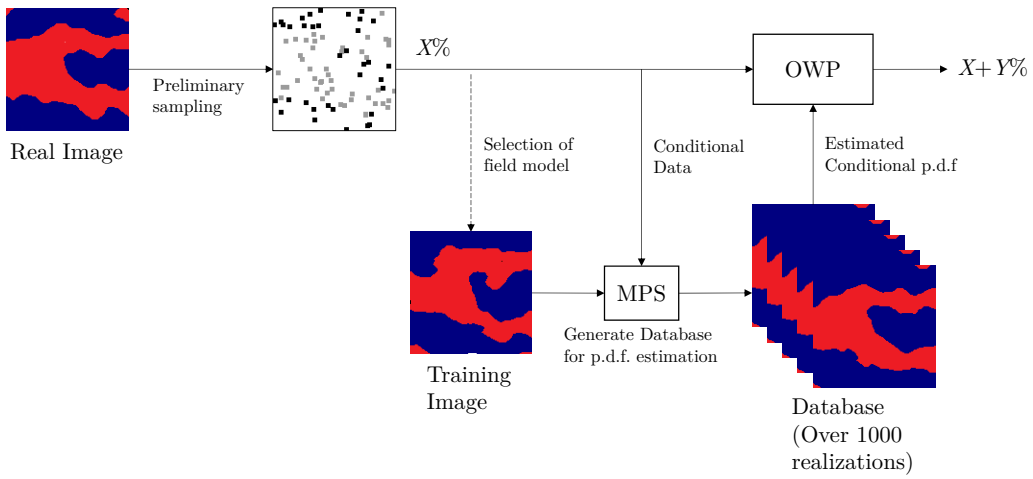

Figure 3.2: Proposed Inference system.

3.3.2 Data Base

The Database used in our preliminary experiments considered the next constrains:

- Nature → Discrete random field
- Dimensionality → *2-D*
- Variable → Permeability
- Dimensions → 200×200 px
- Type → Bi-categorical
- Source → synthetic realizations

3.3.3 Experiments

We considered the next approach to evaluate the sampling design using the proposed database:

- Iterative *OWP* version.
- Combined sampling scheme (uniform sampling (structured) + *OWP*) with an initial presampled positions.
- 200 *MPS* realizations for statistics estimation.

For example, from database we select the next image as the realization image describing the actual channels structure in the subsurface:

Figure 3.3: Realization Image.

First, we used an initial uniform structured sampling [3]and then we applied *OWP* approach. We overall sampling consisted of the 2% samples from the available positions in the field (realization image). Our modified *OWP* sampling take into account the next heuristics rules:

[3]Needed to the right representation of bad conditioned zones, required for the use of *MPS* as a posterior inference system

- Selection of closest conditionals for **each available position for sampling** (9 nearest point in preliminary heuristic implementation).
- Estimation of conditional probabilities by histograms for *MPS* simluations.
- Estimation of entropy maps.
- Selection of maximal entropy candidates and heuristic corrections.

3.3.4 Experimental Results for OWP

Our experimental setting considered binary images that represent permeability fields. The experiments consisted on the comparison between an equally spaced sensing system and a mixed approach that includes equally spaced measurements and measurements given by the iterative algorithms proposed in eq. (3.19). The objective of this comparison was to contrast the effect that sensing design introduces on the *posteriori* entropy map of the field X. For that, simulations *MPS* tool is also evaluated (by SNESIM algorithm)[36]. The *MPS* technique utilizes a training image, provided by expert knowledge, to perform simulated images that conserve the patterns present in the reference image. In addition, the *MPS* method can be conditioned by measured data.

The first approach consisted on an equally spaced samples of the actual realization field. The second one is described in Fig. 3.2. It required a small portion of preliminary data given by the first method and the other portion sequentially obtained by the application of Eq.(3.8). Conditional probability *PDF*s were estimated using the representative database generated from the simulation algorithm [36, 24], which uses *MPS*. Using a frequentist approach, the probability that current pixel takes a specific value is estimated from the simulated data, for each value of the alphabet $\mathcal{A}$ (in this case, $|\mathcal{A}| = 2$). Doing this operation for each position within f^c, we obtained an estimation of:

$$\mathbb{P}(X_{\{\mathrm{i}\}}|X_f), \quad \forall \mathrm{i} \in f^c \tag{3.33}$$

Here, we computed the conditional entropy, required to apply the *OWP* algorithm proposed in Eq.(3.8). The analysis of the results relies on the behavior that the *posteriori* entropy takes as we observe in Fig. 3.4. In this case, white pixels represents a high uncertainty zones, while black pixels means a complete knowledge of the value that the pixel is taking. As expected, for the scheme in which *OWP* is included, the uncertainty is significantly reduced with respect to the other method. Both methods use the same amount of measurements (2% of all $N = 200 \cdot 200$ pixels), but in the one presented in this proposal, we used $X\%$ uniformly sampled and $Y\%$ by our *OWP* method, where $X + Y = 2\%$.

Next step was to generate simulations of the media using *MPS* tool and using the samples provided by the various methods explained above. Using the samples acquired by different sampling schemes, we explored if there is a regime that combines equally spaced data sampling with data sampled through our method, that is near optimal from the point of view of entropy reduction, average reconstruction error, or variability of the

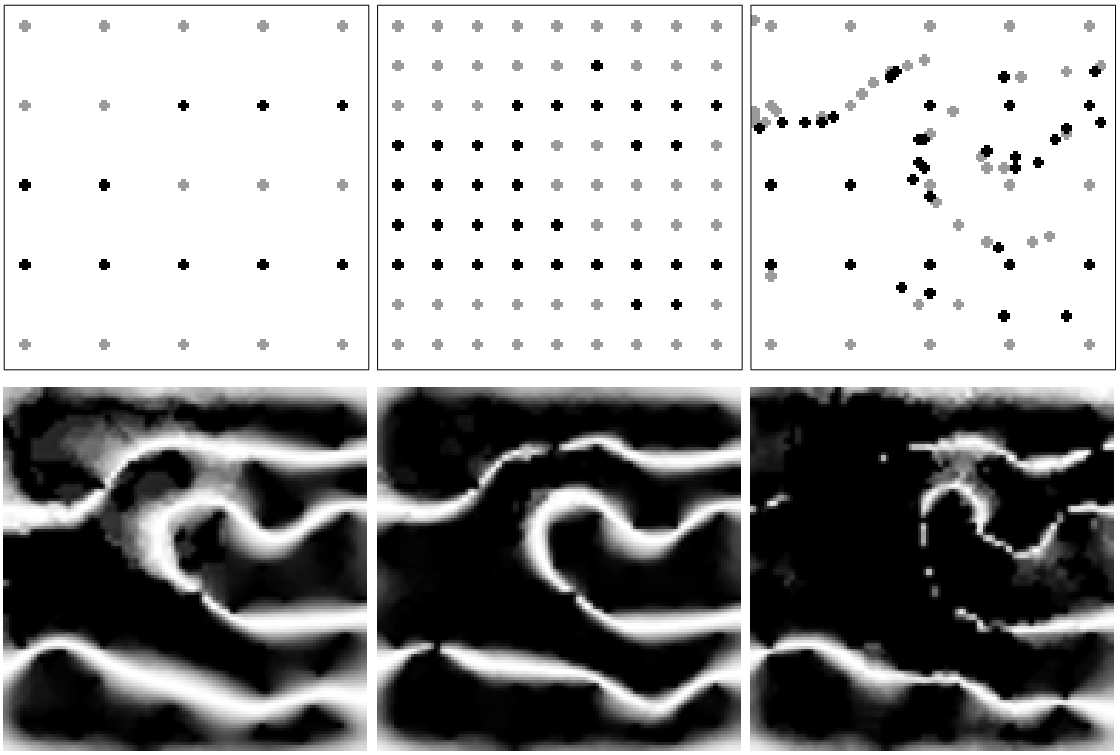

Figure 3.4: Both measurement systems are shown with their *posteriori* entropy map respectively. In the left column partial measurement process (25 uniformly sampled measurements) is shown previous to the *OWP* application. In the middle one, equally spaced sampling method is presented for the 2% of samples. Finally, in the right column, the result after applying *OWP* measurements is shown.

MPS simulations. Two metrics were proposed to measure performance: Signal-to-Noise ratio (SNR) between simulations and the reference image in order to analyze the quality of the realizations; and standard deviation of simulations, to measure the uncertainty introduced.

Table 3.1: $|\sigma|$ corresponds to the total energy present in the standard deviation map. <SNR> represents the mean signal to noise (reconstruction error) ratio between simulations and the realization image. The percentages shown in the first row make reference to the portion of equally spaced data used by the mixed method.

Metric	0%	20%	40%	60%	80%	100%
<SNR>	13.9218	20.0376	23.2711	28.1413	22.9547	28.1044
$\|\sigma\|$	18.2311	17.8426	13.0801	11.1337	10.3974	11.605

We performed 200 simulations for each sampling scheme. As indicated before, we considered the scenario where 2% of samples are available to do the reconstructions based on simulations. The mixed scheme considers 0%, 20%, 40%, 60%, 80% and 100% of equally spaced data, and the results are shown in Table 3.1. From a <SNR> point of view, the optimal case is obtained when combining 60% of equally spaced data and 40% acquired

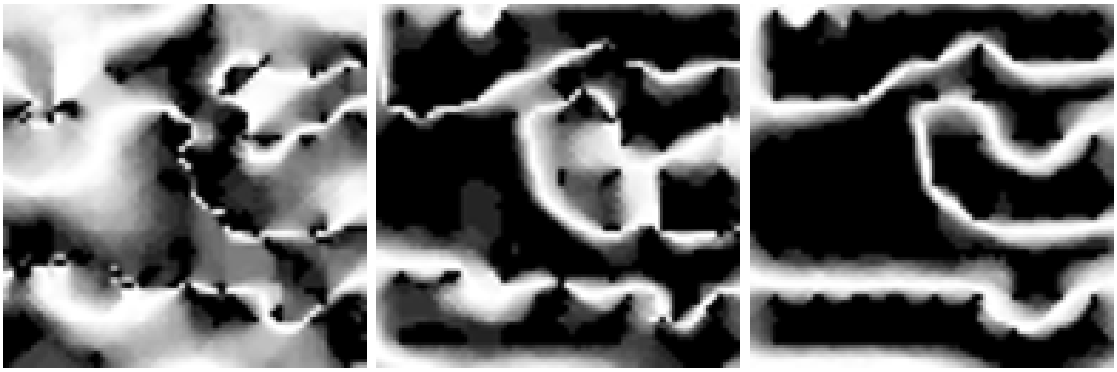

Figure 3.5: Standard deviation maps for different sampling schemes. From left to right we have: full OWP, 60% equally spaced sampling and 40% OWP (optimal from SNR point of view) and a full equally spaced sampling scenario.

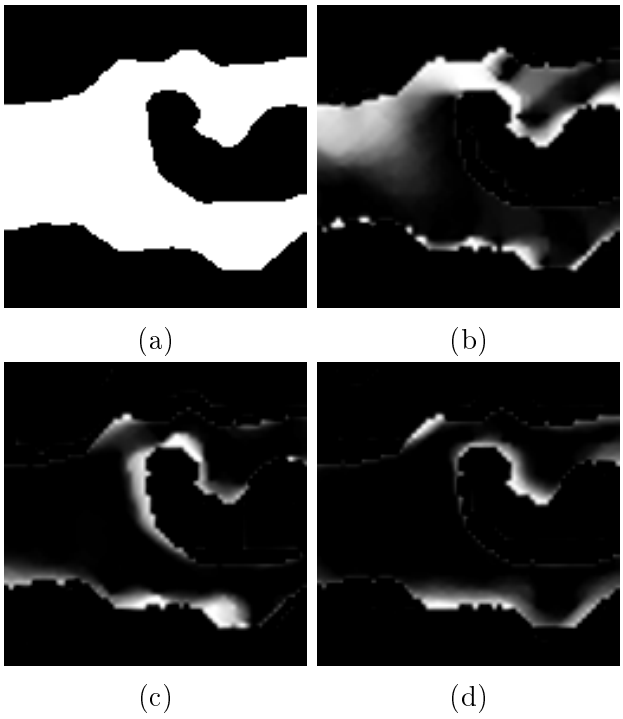

Figure 3.6: a) Reference Image. b) Mean absolute error of full OWP scheme. c) Mean absolute error of 60% equally spaced and 40% OWP scheme. d) Mean absolute error of full equally spaced scheme.

through OWP method (see Fig. 3.6). For the standard deviation case, optimum is placed in the range of the $80-20\%$ combination (see Fig. 3.5). These results indicated that both kind of data placement methods are complementary and that there exists a trade-off between characterization and uncertainty reduction. Provided results were coherent with the proposed algorithm nature that attempt to reduce global uncertainty on the overall process.

3.4 On Noisy Compressive Sensing Performance

NCS theory has motivated us to study the spatial co-dependencies of the regionalized variable of interest. heretofore, we have investigated in how *MPS* can help in achieving this task and we have initiated an analysis of how the incorporation of spatial dependence improves sparse promoting algorithms.

3.4.1 Statistical Analysis from *MPS*

Using the data base of multichannel MC_1 model we obtained 200 *MPS* realizations from several hard data measurements levels. For each hard data regime, then we estimated second order statistics. As shown in fig. 3.7 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 hard data measurements increased a reduction of non local variance was observed, thereby the powerful of spatial conditioning from *MPS* was validated.

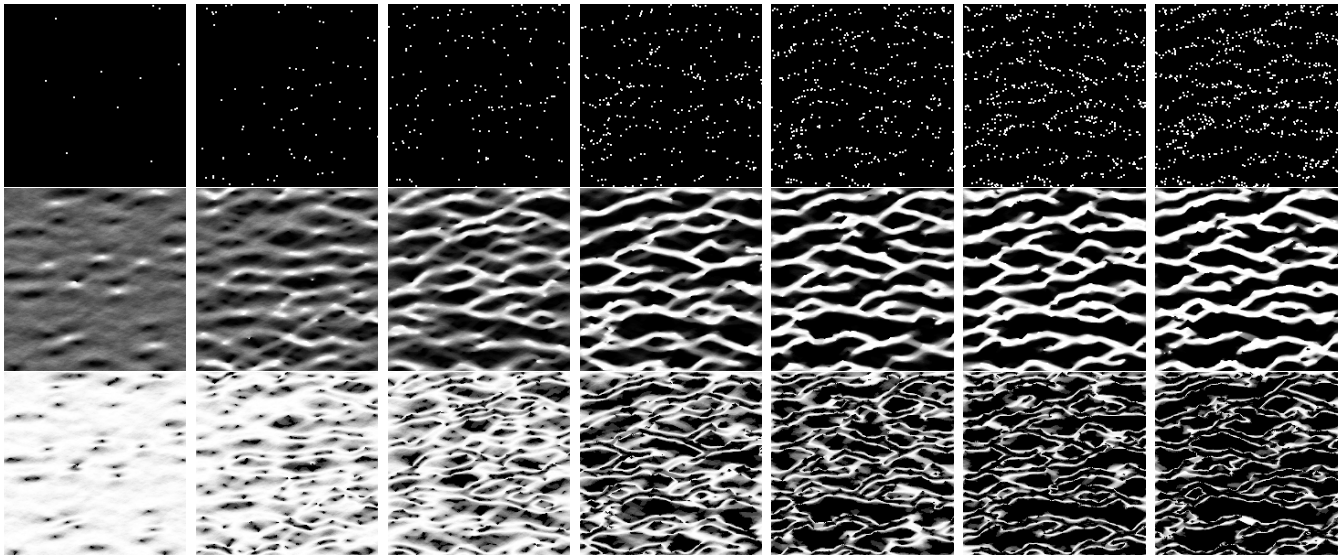

Figure 3.7: Statistics for simulations from 0.1 %, 0.5 %, 1 %, 2 %, 3 %, 4 %, 5 % hard data scenarios

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 let us to propose the use of *MPS* realizations as an estimation of the spatial covariance between regionalized variables conforming the stochastic field.

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

3.4.2 *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 withe 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 hard data measurements are considered as input to the reconstruction methods and no one knowledge is given for unmeasured regionalized variables. The incorporation of *MPS* allows no 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. (2.15) but with $C_v^{\frac{-1}{2}}$ 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.

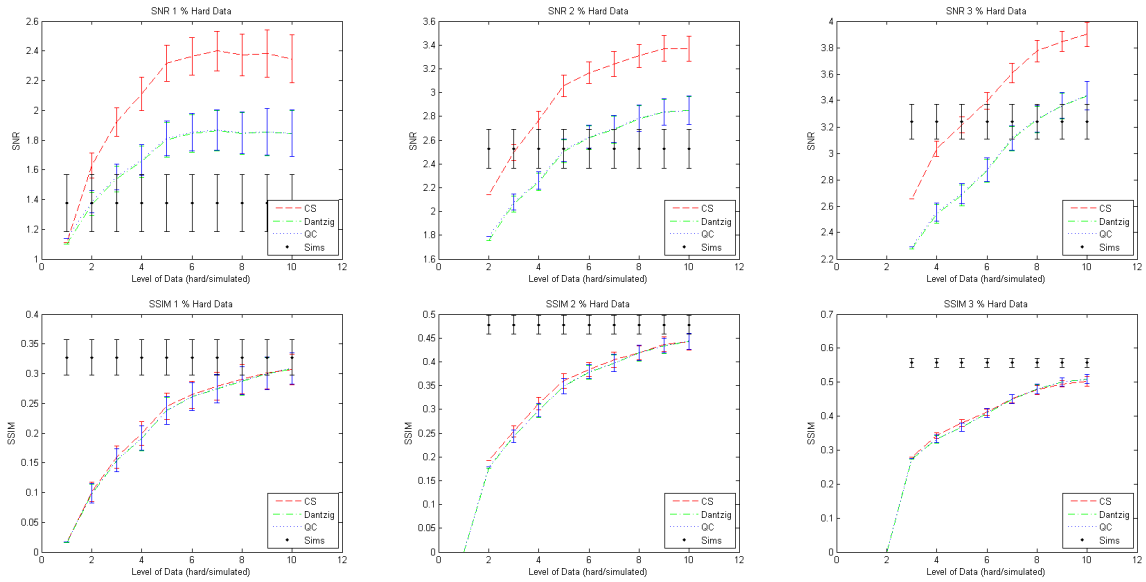

Figure 3.8: Performance Analysis of Naive Approach of *NCS*by only relaxing restrictions on soft data. From an initial amount of hard data measurements some level of soft virtual measurements are added from simulations and then *CS* and naive *NCS* approaches are applied. Upper row *SNR* analysis, lower 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. 3.8, 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* achieve a better global performance.

Without incorporating spatial correlations on the regionalized variables we are not taken advantages from *NCS* theory. Including uncertainty on measures only increase the searching space for the optimization algorithms. Thus, our next motivation was including additional constraints promoting desired features on the regionalized variables.

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

In figure 3.9 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. 3.10.

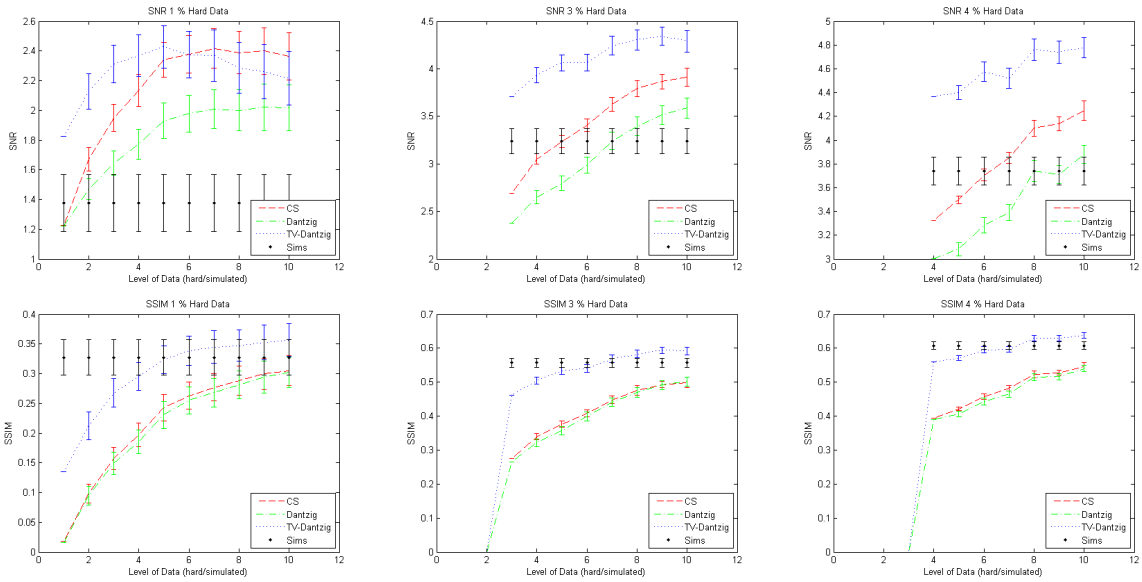

Figure 3.10: Performance Analysis of Naive Approach of *NCS*by only relaxing restrictions on soft data adding TV as sparsity promoting approach. From an initial amount of hard data measurements some level of soft virtual measurements are added from simulations and then *CS* and naive *NCS* approaches 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.

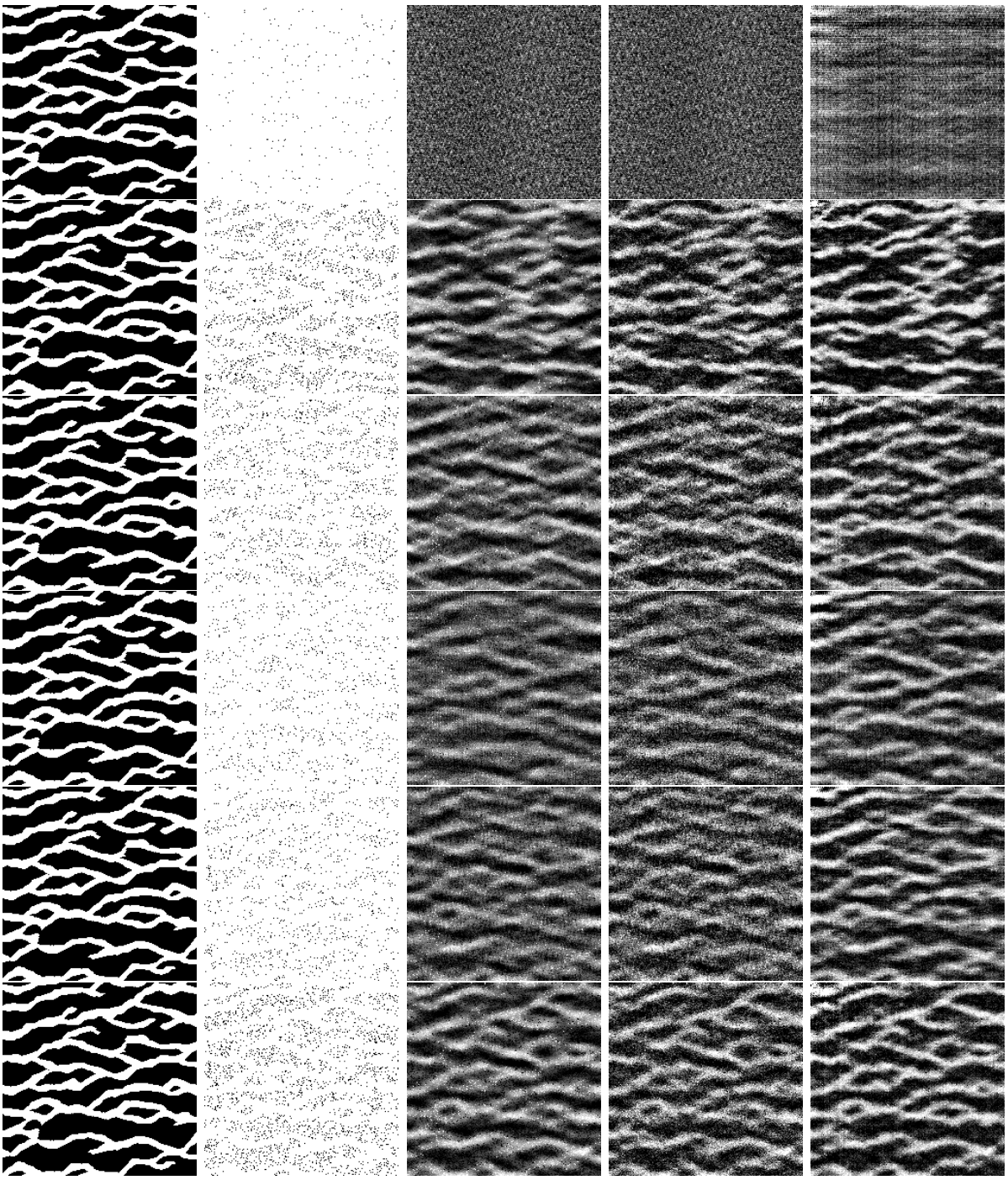

Figure 3.9: Examples of outcomes for *NCS* without considering spatial dependence. From left to right: Target field image, hard data plus simulated data (HD+SD level), standard *CS* reconstruction for HD+SD, *NCS* for HD+SD by Dantzig selector approach, *NCS* for HD+SD by Dantzig selector and *TV* approach. 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*

3.4.3 *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 we have only qualitative results which account for an apparent improvement in the method's ability to capture something of the structure of original image (figs. 3.11, 3.12, 3.13 and 3.14).

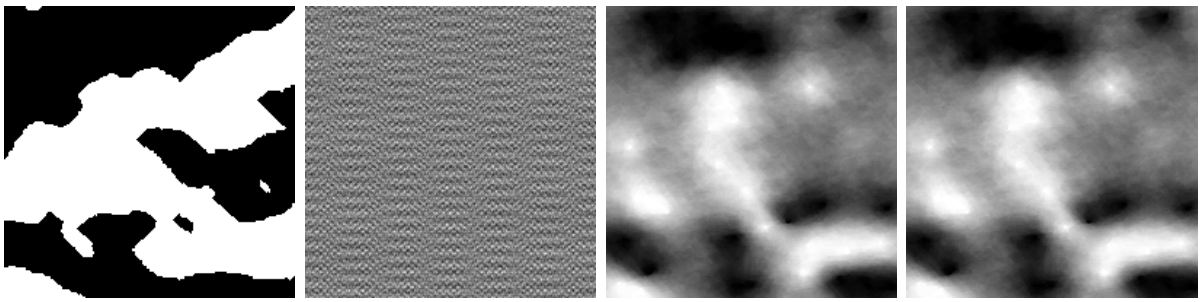

Figure 3.11: Example of *NCS* reconstruction under assumption of independence. Single Channel 1: Hardata 0.1 %. True Image. Standard *CS*. *NCS* by quadratic constraints. NCS by Dantzig selector.

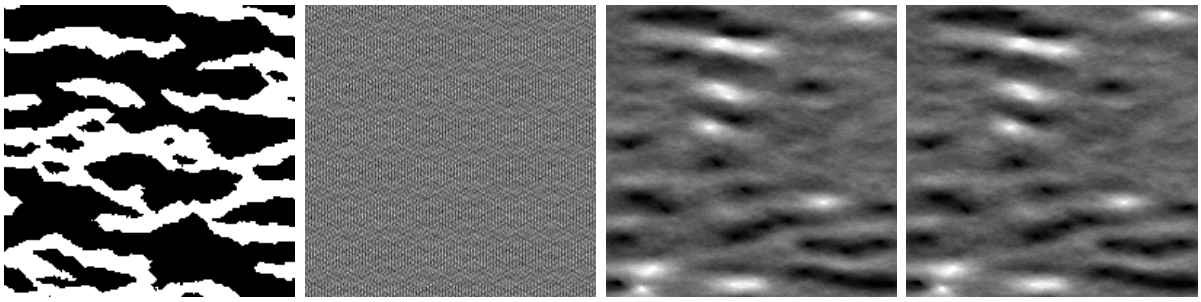

Figure 3.12: Example of *NCS* reconstruction under assumption of independence. Multi Channel 1: Hardata 0.1 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig.

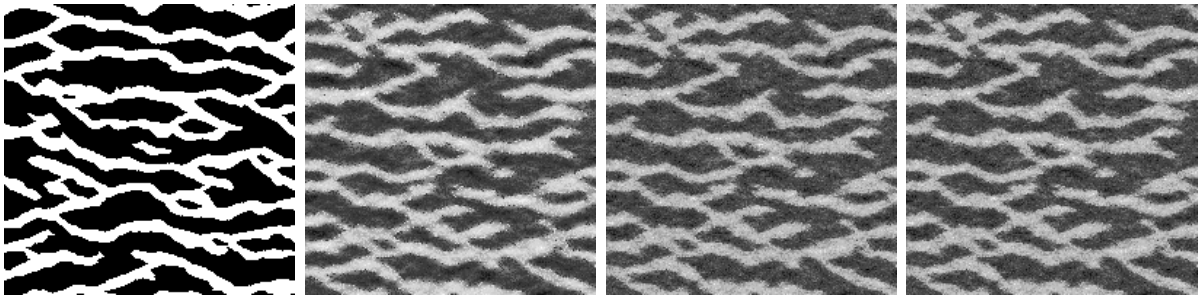

Figure 3.13: Example of *NCS* reconstruction under assumption of independence. Multi Channel 2: Hardata 0.1 %, Soft data 29.9 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.

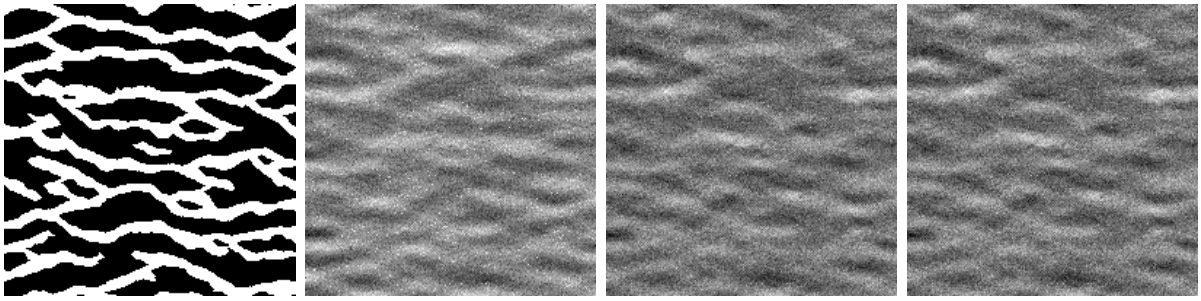

Figure 3.14: Example of *NCS* reconstruction under assumption of independence. Multi Channel 2: Hardata 0.1 %, Soft data 3.9 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.

3.4.4 *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 we have only qualitative results which account for an apparent improvement in the method's ability to capture both global structure and deltails of original image from low sampling regimes (figs. 3.15 and 3.16).

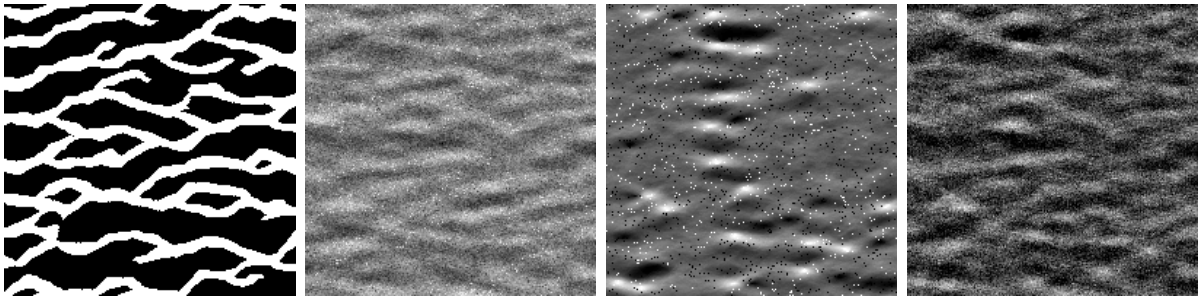

Figure 3.15: Example of *NCS* with full covariance estimation. Multi Channel 2: Hardata 0.1 %, Soft data 3.9 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.

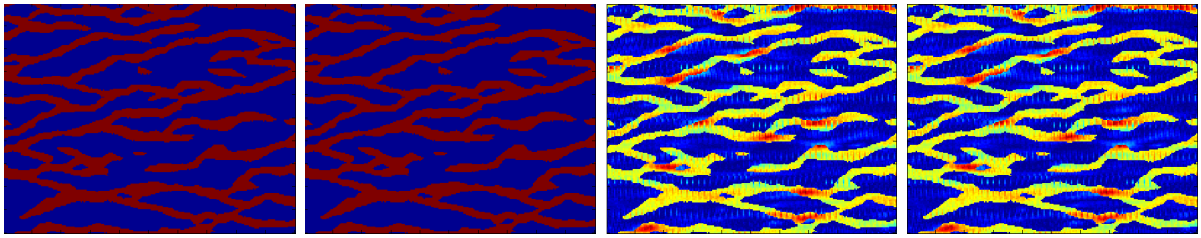

Figure 3.16: Example of *NCS* with full covariance estimation. Multi Channel 2: Hardata 1 %, Soft data 99 %. True Image. Standard CS. NCS by quadratic constraints. NCS by Dantzig selector.