SENSITIVITY ANALYSIS OF FILTERSIM AND HISTOGRAM REPRODUCTION

A REPORT
SUBMITTED TO THE DEPARTMENT OF ENERGY RESOURCES ENGINEERING
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

By
Bruno Dujardin
June 2007
I certify that I have read this report and that in my opinion it is fully adequate, in scope and in quality, as partial fulfillment of the degree of Master of Science in Petroleum Engineering.

Andre Journel
(Principal advisor)
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>v</td>
</tr>
<tr>
<td>Preface</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td><strong>1 Sensitivity analysis of filtersim</strong></td>
<td>6</td>
</tr>
<tr>
<td>1.1 2D-2facies sensitivity analysis</td>
<td>6</td>
</tr>
<tr>
<td>1.2 2D sensitivity analysis with continuous variables</td>
<td>16</td>
</tr>
<tr>
<td>1.3 Conclusions</td>
<td>28</td>
</tr>
<tr>
<td><strong>2 Reproduction of histogram with filtersim</strong></td>
<td>29</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>29</td>
</tr>
<tr>
<td>2.2 Histogram reproduction</td>
<td>31</td>
</tr>
<tr>
<td>2.3 Conclusion and future work</td>
<td>47</td>
</tr>
<tr>
<td><strong>3 Impact of mps on estimation</strong></td>
<td>48</td>
</tr>
<tr>
<td>3.1 Two facies simulation</td>
<td>48</td>
</tr>
<tr>
<td>3.2 Four facies simulation</td>
<td>64</td>
</tr>
<tr>
<td>3.3 Conclusion</td>
<td>66</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td>75</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Default parameters ................................................................. 31
2.2 Sensitivity of spatial variance and maximum value to number of realizations 33
2.3 Time for classification + realizations ........................................ 37
2.4 Time for classification at different grids .................................... 37
2.5 Time for classification and realization ....................................... 42
2.6 Reproduction of basic statistics ................................................. 43
2.7 Measure of ergodic fluctuations ................................................ 44
2.8 Measure of horizontal effective permeability .............................. 46
3.1 Measure of local accuracy ....................................................... 66
# List of Figures

1. 2D continuous training image ................................................................. 4
2. Histogram of mean and variance of 100 filtersim realizations run with default parameters ......................................................... 4

1.1 2D-2facies training image ................................................................. 7
1.2 Three realizations with default parameters on a 100 × 100 grid ............ 7
1.3 Realizations with different cross partition numbers (sand target: 28%) .... 9
1.4 Time for one realization as a function of the Initialization number of the cross partition. The total number of bins (=max number of parents) is equal to Initialization#offilters ................................................................. 9
1.5 Realizations with different K-mean partition numbers (sand target: 28%) . 10
1.6 Time for one realization as a function of the Initialization number of the K-mean partition ................................................................. 10
1.7 Template input .................................................................................. 11
1.8 Realizations for different template inputs .............................................. 12
1.9 Realizations for different number of multigrids ...................................... 13
1.10 Realizations for different numbers of Cmin ........................................... 14
1.11 Realizations for different numbers of servosystem .................................. 15
1.12 Three realization for each partition type and frozen parameters values .... 15
1.13 250 × 250 continuous training image and its histogram ....................... 17
1.14 Realizations with cross partition (Initialization number increasing from 2 to 10) Row 2: Q-Q plots with the training image Row 3: Histogram of the realizations ................................................................. 18
1.15 Time for one realization as function of the Initialization number of the Cross partition ................................................................. 18
1.16 Superposition of one realization (cross partition with a Initialization number=3) over the training image

1.17 Realizations with K-mean partition (Initialization number increasing from 100 to 1000) Row 2: Histogram of the realizations

1.18 Time for one realization as function of the Initialization number of the K-mean partition

1.19 Realizations for different template parameter inputs

1.20 Realizations for different number of multgrids with default parameters

1.21 Realizations for different number of Cmin with default parameters

1.22 Realizations for different servosystem number with default parameters

1.23 E-type of 100 non conditional realizations for a 100*100 and a 250*250 grid

1.24 Spatial variance fluctuations over 20 non conditional realizations for a 100*100 and a 250*250 grid

2.1 Histogram of 100 realizations performed with new method

2.2 2D continuous training image

2.3 Default parameters

2.4 Realizations with default parameters

2.5 Classification at fine grid

2.6 Histogram of mean of prototypes for K-mean partition at the first grid

2.7 Different steps of the simulation process

2.8 Realizations with histogram transformation

2.9 Histogram of realizations with histogram transformation

2.10 Realizations with score distance

2.11 Realizations with histogram transformation + distance score

2.12 Histogram of realizations with histogram transformation + distance score
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.13</td>
<td>Different steps of the simulation process</td>
<td>39</td>
</tr>
<tr>
<td>2.14</td>
<td>Realizations on 250 × 250</td>
<td>40</td>
</tr>
<tr>
<td>2.15</td>
<td>250 × 250 simulation grid</td>
<td>41</td>
</tr>
<tr>
<td>2.16</td>
<td>Miror TI (500 × 500)</td>
<td>41</td>
</tr>
<tr>
<td>2.17</td>
<td>Realizations with larger TI</td>
<td>42</td>
</tr>
<tr>
<td>2.18</td>
<td>Histogram of realizations with larger TI</td>
<td>42</td>
</tr>
<tr>
<td>2.19</td>
<td>TI variogram map (lag of 125)</td>
<td>45</td>
</tr>
<tr>
<td>2.20</td>
<td>Variogram maps for 100 realizations (lag of 50)</td>
<td>46</td>
</tr>
<tr>
<td>3.1</td>
<td>Reference data set</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Training image</td>
<td>51</td>
</tr>
<tr>
<td>3.3</td>
<td>Kriging</td>
<td>52</td>
</tr>
<tr>
<td>3.4</td>
<td>Sisim</td>
<td>53</td>
</tr>
<tr>
<td>3.5</td>
<td>Filtersim</td>
<td>55</td>
</tr>
<tr>
<td>3.6</td>
<td>Indicator maps ensuring the same global proportion of sand</td>
<td>57</td>
</tr>
<tr>
<td>3.7</td>
<td>Connectivity measure for one sisim realization (blue), one filtersim realization (red) and the reference image (dark)</td>
<td>59</td>
</tr>
<tr>
<td>3.8</td>
<td>Binary images after thresholding at 0.4 the kriging estimate, the sisim and filtersim E-type maps</td>
<td>60</td>
</tr>
<tr>
<td>3.9</td>
<td>EW bandwidth connectivity (bandwidth=15)</td>
<td>61</td>
</tr>
<tr>
<td>3.10</td>
<td>Variograms in EW and NS directions</td>
<td>63</td>
</tr>
<tr>
<td>3.11</td>
<td>Data set</td>
<td>67</td>
</tr>
<tr>
<td>3.12</td>
<td>Indicator kriging maps</td>
<td>68</td>
</tr>
<tr>
<td>3.13</td>
<td>Histograms of indicator kriging maps</td>
<td>69</td>
</tr>
<tr>
<td>3.14</td>
<td>Binary images after thresholding the indicator kriging maps</td>
<td>70</td>
</tr>
<tr>
<td>3.15</td>
<td>Sisim and filtersim realizations</td>
<td>71</td>
</tr>
<tr>
<td>3.16</td>
<td>Filtersim E-type on 200 realizations</td>
<td>72</td>
</tr>
<tr>
<td>3.17</td>
<td>Binary images after thresholding filtersim E-type maps, ensuring target proportions</td>
<td>73</td>
</tr>
<tr>
<td>3.18</td>
<td>EW connectivity for the reference (black), the filtersim E-type (red), the kriging estimate (green)</td>
<td>74</td>
</tr>
</tbody>
</table>
Preface

Multiple-point statistics (mps) algorithms generate reservoir models that aim at reproducing geological patterns taken from a training image. These algorithms were implemented first in the snesim code developed by Strebelle (2000) then in the filtersim code developed by Zhang et al. (2006). While snesim is limited to facies categorical simulation, filtersim can handle both continuous and categorical variables. Though these algorithms have shown success in providing good reproduction of geological patterns present in hydrocarbon reservoirs, they still need to be improved to integrate more complex geology. As there is no unique mps algorithm, each algorithm was fully tested to understand how to tune its parameters. This work presents such extensive sensitivity analysis.

In petroleum applications, simulated reservoir models are used in flow simulations that are very sensitive to the connectivity of low or high permeability regions. So it is important to control the reproduction of such connectivity patterns that can be either barriers or channels to flow. A sensitivity analysis is performed on the main parameters of filtersim to provide some guidelines on how to use them. Based on this analysis, this thesis proposes a method to improve the reproduction of the training image histogram.
Introduction

The filtersim algorithm starts by classifying training patterns extracted from a training image: a set of filters is applied on each training pattern to reduce it to a small number of score values, these patterns are then grouped within prototype classes defined in the filter score space according to score similarities. Simulation proceeds by visiting the nodes of the grid along a random path. At each node, the local conditioning data event is compared to all prototype classes to find the closest class, a pattern from that closest class is drawn and is pasted onto the simulation grid. To capture the large scale structures of the training image, the simulation process is performed over multi-grids from coarse to fine using at each grid the information from the previously simulated coarser grid. This stochastic simulation process allows to generate different realizations all conditioned to the same set of data. In each realization, reproduction of the training image patterns and of the target histogram need to be tested. The reproduction of the training image target histogram is not ensured by the original filtersim algorithm and calls for a specific processing.

This thesis built on the multiple-point geostatistical simulator filtersim consists of three major parts: (1) Sensitivity analysis of filtersim, (2) Reproduction of histogram with filtersim and (3) Impact of multiple-point statistics on estimation, as opposed to simulation.

Sensitivity analysis of filtersim

Sensitivity analysis is performed on the important input parameters of the filtersim program (Zhang et al. (2006)) as coded by Wu (2007) into the freeware SGEMS (Stanford Geostatistical Earth Modeling Software). The results are analyzed and guidelines are provided on how to set and use these parameters. In filtersim, parameters are accessed through four menus: General, Conditioning, Region and Advanced. The General menu defines the
training image, the template dimension and the data type. The Conditioning menu is used to enter the hard and/or soft data, the Region menu defines the different regions to be simulated. The Advanced menu defines the parameters used for pattern classification. In this report, a sensitivity analysis is done using first a 2D-2facies categorical training image, then using a continuous training image. No conditioning data and a single region were retained; therefore only parameters from the General and the Advanced Conditioning menu are required. The quality of realizations and the CPU cost depend on the sequence of processing steps in filtersim: Filters $\rightarrow$ Classification $\rightarrow$ Simulation process. For a given set of default filters, classification is the most sensitive parameter. The original servo-system that aims at reproducing the target proportion does not work properly.

Reproduction of histogram with filtersim

The multiple-point simulation program filtersim generates simulated realizations that aim at reproducing the geological features of the input training image. Such geological features involve structural statistics much beyond the traditional 2-points statistics or variograms. However reproduction of a target histogram and especially its mean and variance, which are one-point statistics is still lacking. Consider, for example, the 2D continuous $250 \times 250$ training image given on Figure 1(a). It has a mean of 1.93, a variance of 1.06 and its values are in the $[0.07, 7.15]$ interval. Its histogram given in Figure 1(b) shows a mode around 0.7 and a long tail of high values. This training image is used to perform unconditional simulation on a $100 \times 100$ grid with filtersim.

Figures 2(a) and 2(b) give the histogram of the means and the variances of 100 filtersim realizations run with the program default parameters. These distributions are not centered on the target mean and variance set equal to the training image (mean $= 1.93$ and variance $= 1.06$). All 100 realizations underestimate significantly the target variance of the training image. To correct this bias, we suggest removing the original servosystem and replace it by an algorithm based on a transform aimed at reproducing not only the target mean or proportions, but also an entire target histogram including its variance and proportions of extremes values.
(a) Training image  

(b) Training image histogram

Figure 1: 2D continuous training image

(a) Histogram of mean (target: 1.93)  

(b) Histogram of variance (target: 1.06)

Figure 2: Histogram of mean and variance of 100 filtersim realizations run with default parameters
Impact of multiple-point statistics on estimation

Once the filtersim algorithm is properly calibrated, it can be applied to assess local uncertainty. The filtersim E-type estimate defined as the pointwise average of many simulated realizations is compared to that based on traditional 2-point simulations (program sisim, Sequential Indicator SIMulation) and to the direct kriging estimate. The comparison is done in terms of local accuracy and reproduction of relevant structural statistics such as a facies connectivity measure. This study is performed with a 2 facies training image then with a 4 facies training image. The sensitivity of the estimation (E-type or kriging) to the facies geometry is analysed. The filtersim E-type is slightly better than the direct kriging estimate both in terms of local accuracy and reproduction of structural statistics of the training image.
Chapter 1

Sensitivity analysis of filtersim

1.1 2D-2facies sensitivity analysis

Figure 1.1 gives a 2D training image of size $250 \times 250$ with 2 facies depicting sand channel in a mud background. This training image is retained to simulate non-conditional realizations on a smaller $100 \times 100$ grid. The target marginal distribution is set to the training image proportions: 72% mud and 28% sand. The default weights parameters: 0.5, 0.3, 0.2 are retained for calculation of the distance used to determine the class prototype closest to the conditioning data event. The first weight is for hard data, the second one for simulated data and the third one for data carried by patching. The default seed number 211175 is used. From one simulation to another one, changing the seed number does not change the quality of the realizations. Figure 1.2 shows three realizations obtained using different seed numbers with all parameters set to their default values. The following parameters, as input into the General and Advanced menu, are analyzed in the order they are presented. The values in parentheses represent the default parameter values used unless specified otherwise:

- Partition type with Initialization and Secondary numbers (Cross partition with Initialization = 2 and Secondary = 2, see definitions later)
- Template dimension: Search template ($11 \times 11 \times 1$) and Inner Patch ($7 \times 7 \times 1$) dimensions
- Number of Multigrids (3)
- Cmin of each grid (10)
1.1. 2D-2FACIES SENSITIVITY ANALYSIS

- Servosystem factor (0.5)

The Partition type is the first parameter studied because it has been found to be the most critical parameter from both a CPU and a quality of results points of view.

![2D-2facies training image](image1)

Figure 1.1: 2D-2facies training image

![Three realizations with default parameters on a 100 × 100 grid](image2)

Figure 1.2: Three realizations with default parameters on a 100 × 100 grid

1.1.1 Partition type

In filtersim, the partition of the filter score space is either done by cross partition or by partition with K-mean. The partition type and its parameters, especially the Initialization number are the most critical parameters to get good and fast realizations. For each partition type, there are an Initialization number for the first partition of the score space and a Secondary number for secondary partitions. In the first partition, the Initialization number defines the number of bins for the classification of each filter score. The Secondary number
defines for all following partitions the maximum number of subdivisions of each bin. For instance, in the Cross partition type with 6 filters and an Initialization number of 2, there are a maximum of $2^6 = 64$ bins for the first partition, with most of these bins empty which reduces the number of actual bins. Then, each non-empty bin can be further partitioned into a maximum of 3 bins if the Secondary number is 2. These first children bins can be further partitioned, each into up to 3 bins. In the K-mean partition option, the single Initialization number defines the final number of bins (parents with no children) used for partitioning the score space.

**Cross partition**

By increasing the Initialization number, realizations are improved as shown in Figure 1.3, but at greater CPU cost as shown in Figure 1.4. The secondary number is less critical and less time consuming because it results in only small modifications of the data structure (a link of link) where prototypes are stored. On Figure 1.3, increasing the secondary number while keeping constant the Initialization number does not improve the realizations in terms of channel continuity and sand proportion reproduction. On Figure 1.3, the red box shows the realization with default values (2,2) for the Initialization and the Secondary numbers: the result is poor in that the channel continuity is not reproduced and the proportion of sand is only 21% instead of the target 28%. With an Initialization number of 4, the realization is good but it is 50% more time consuming. It is not efficient to increase the Initialization number beyond 4 or 5. The right graph of Figure 1.4 shows that the CPU is proportional to the number of bins of the first partition (run time are given for a Pentium 4, CPU 3.0GHz). Thus the choice of a correct value for that Initialization number is decisive for CPU time.

**K-mean partition**

By increasing the Initialization number of the K-mean partition, realizations are improved as seen on Figure 1.5, but at greater CPU cost as shown on Figure 1.6. To get good realizations for this example, the Initialization number has to be higher than 2000 and it takes 35 times more CPU than the cross partition option with an Initialization number of 5 (15625 bins in the first partition), see Figure 1.3. In the following, an Initialization number of 3 and a Secondary number of 2 for the Cross partition method are kept constant (see the
1.1. 2D-2FACIES SENSITIVITY ANALYSIS

Figure 1.3: Realizations with different cross partition numbers (sand target: 28%)

Figure 1.4: Time for one realization as a function of the Initialization number of the cross partition. The total number of bins (=max number of parents) is equal to $Initialization^{\# of filters}$.
CHAPTER 1. SENSITIVITY ANALYSIS OF FILTERSIM

Figure 1.5: Realizations with different K-mean partition numbers (sand target: 28%)

Figure 1.6: Time for one realization as a function of the Initialization number of the K-mean partition

green box on Figure 1.3). These parameter values provide realizations that are not too far from the target statistics but still need to be improved. Beware that these frozen parameter values would be different if a different training image was considered. The richer (more varied) and the larger the training image, the higher the Initialization number needed to keep the same ratio between parents (issued from the initial partition) and children (issued from secondary partitions). The ratio between the number of children and the number of parents could be a good indicator to characterize a partition.
1.1.2 Template dimension

Frozen parameters: Cross partition with Initialization=3 and Secondary=2

The template input is constituted of a Search template part and an Inner patch part (Figure 1.7). The inner patch represents the part of the template that is frozen on the simulated field once a training pattern has been drawn. The dimensions given hereafter are for the finest grid of a multiple-grid setting. With a $250 \times 250$ training image grid and a $100 \times 100$ simulation grid, the default search template dimension of $11 \times 11$ provides good results (4th column of Figure 1.8). The best results are provided when the dimensions of the search template and the inner patch are in reasonable ratio: respectively $(11 \times 11$ and $7 \times 7)$ or $(15 \times 15$ and $11 \times 11)$. Decreasing too much the inner patch dimension for a given search template dimension does not improve the realization and is time consuming. Indeed on the $(11 \times 11$ and $3 \times 3)$ realization (Figure 1.8), there appear fewer channels than in the $(11 \times 11$ and $7 \times 7)$ realization. So a reasonable ratio of Search template vs Inner patch should be kept. The choice of the template dimension should depend on the sizes of the simulation grid and of the training image but also on the number of multigrids. Indeed with more multigrids, choosing a smaller template dimension (at the finest scale) allows to scan the same scale as a bigger template dimension with fewer multigrids.

![Figure 1.7: Template input](image)
CHAPTER 1. SENSITIVITY ANALYSIS OF FILTERSIM

Figure 1.8: Realizations for different template inputs
1.1.3 Number of multigrids

Frozen parameters: Cross partition with Initialization=3 and Secondary=2, search template= $11 \times 11$ and inner patch= $7 \times 7$.

A multigrid simulation, using a template of greater size for the coarser grid, allows capturing the large scale structures of the training image. For instance an $11 \times 11$ Search template parameter at the finest grid represents a $41 \times 41$ Search template at a third coarse grid. With the number of multigrids set to one, the simulation is done directly on the final grid. On Figure 1.9, the single grid realization does not reproduce well the channel structure of the training image and the sand proportion of 39% is too high (target is 28%). By increasing the number of multigrid to 2 then 3, the simulated channels of the realizations match better those of the training image. Increasing the number of multigrid above 4 does not change the simulation because the maximum number of multigrids is automatically controlled by the respective sizes of the training image and simulation grid. In the case of a $250 \times 250$ training image with a $100 \times 100$ simulation grid, the default number of multigrids of 3 is a good number.

![Figure 1.9: Realizations for different number of multigrids](image)

1.1.4 Cmin to control the partition process

Frozen parameters: Cross partition with Initialization=3 and Secondary=2, search template= $11 \times 11$, inner patch= $7 \times 7$ and 3 multigrids.

During the partition process, secondary partitions are done for those classes whose prototypes are not "sharp" enough and have a number of replicates greater than Cmin, and also on classes that have more than $2 \times \text{Cmin}$ replicates. By decreasing Cmin, the partition becomes more refined and simulations are improved a bit. On Figure 1.10 with Cmin =...
5, channels are more continuous and the sand proportion is better reproduced (24%) than from realizations with Cmin = 10 or 20. Decreasing the Cmin does not seem to be time consuming because it does not affect much the final partition. For instance, with an equal number of parents (295), the number of children at the finest grid scale with Cmin=5 is 34,336 instead of 33,499 when Cmin=20.

![Realizations for different numbers of Cmin](image)

Figure 1.10: Realizations for different numbers of Cmin

### 1.1.5 Servosystem

Frozen parameters: Cross partition with Initialization=3 and Secondary=2, search template= 11 × 11, inner patch=7 × 7, 3 multigrids and Cmin=10.

The servosystem parameter aims at reproducing better the target facies proportions in each realization. It is recommended to select a training image which proportions are close to the target proportions. On Figure 1.11, when the servosystem is set lower than 0.95, the proportions in the different realizations fluctuate between 21 and 22% far from the target 28%. Yet, when the servosystem is set too high, the channel continuity could be broken. On Figure 1.11, when the servosystem is set to its maximum 1, the simulated channels are broken; they are somewhat more continuous on the realization with a high servosystem set at 0.95. The servosystem algorithm as presently coded is not satisfactory and need to be improved. This is a major goal of this thesis work.
1.1. 2D-2FACIES SENSITIVITY ANALYSIS

1.1.6 Conclusion

With the 250 × 250 training image of Figure 1.1 and a 100*100 simulation grid, the best compromise between quality and CPU time is obtained for a Cross partition method with an Initialization number = 5 and with all default values for the other parameters (search template=11 × 11, inner patch=7 × 7, 3 multigrids, Cmin=10, servosystem=0.5). To get the same quality with the Kmean partition method, an Initialization number of 2000 is needed, but it takes 35 times more CPU. On Figure 1.12 three realizations for each of these two partition methods are given. By increasing the Initialization number for each of this two partition methods, quality could still be improved but at much greater CPU cost especially for the Kmean partition method.

![Figure 1.12: Three realizations for each partition type and frozen parameters values](image)
1.2 2D sensitivity analysis with continuous variables

Figure 1.13 gives a 2D continuous training image generated starting from the 2D binary training image used in the first part (Figure 1.1). Using a moving average process, this binary image was transformed into a continuous image. Then sequential Gaussian simulation was performed with a sample of data taken from this continuous image to produce a second image. Last, these two previous images were combined to yield the 2D reference continuous image (250 × 250) on Figure 1.13 where channel continuity can still be seen. The distribution of this continuous training image is unimodal (Figure 1.13) with mean=1.91, min=0.14, max=3.95 and variance=0.44. The variable can be interpreted loosely as porosity. Sensitivity analysis with the same sequence and same default parameters as used for the previous categorical case is conducted:

- Partition type with Initialization and Secondary numbers (Cross partition with Initialization = 2 and Secondary = 2)
- Template dimension: Search template (11 × 11 × 1) and Inner Patch (7 × 7 × 1) dimensions
- Number of Multigrids (3)
- Cmin of each grid (10)
- Servosystem factor (0.5)

1.2.1 Partition type

The partition type options for a continuous variable are the same as for the categorical case previously discussed: cross partition or K-mean partition with the same input parameters.

Cross partition

All the realizations of Figure 1.14 reproduce reasonably well the target distribution histogram and there is no improvement by increasing the Initialization number from a distribution point of view. By looking at the Q-Qplot on Figure 1.14 there is no clear difference between realizations. The first realization with an Initialization number=2 is already
satisfactory from a distribution point of view. However, the spatial variances of all these realizations are smaller than that of the training image (0.44). So reproducing the target histogram appears less of a critical problem for continuous filtersim; it is also not a relevant criterion to judge the quality of realizations. When the Initialization number increases, the realizations are smoother and the patch size artifact is less apparent, see the last realization with Initialization number=10. But increasing the Initialization number comes at greater CPU cost (Figure 1.15). On Figure 1.16, by superposing one realization (100 × 100) on the training image (250 × 250), reproduction of the size of the training patterns is checked.

**K-mean partition**

On Figure 1.17, for all Initialization numbers, the histogram reproduction appears good (target mean=1.91 and variance=0.44). By increasing the Initialization number, realizations are improved (Figure 1.17) but at great CPU cost (Figure 1.18). Recall that to get good realizations with the Kmean partition takes 35 times more CPU than with cross partition in the categorical case. In the following, the Initialization number of 3 and the Secondary number of 2 for the Cross partition method are kept constant (see the green box on Figure 1.14). These parameter values provide realizations that are not too far from the target statistics but could still be improved. As before, beware that these frozen parameter values
CHAPTER 1. SENSITIVITY ANALYSIS OF FILTERSIM

Figure 1.14: Realizations with cross partition (Initialization number increasing from 2 to 10)  
Row 2: Q-Q plots with the training image  
Row 3: Histogram of the realizations

Figure 1.15: Time for one realization as function of the Initialization number of the Cross partition
1.2. 2D SENSITIVITY ANALYSIS WITH CONTINUOUS VARIABLES

Figure 1.16: Superposition of one realization (cross partition with a Initialization number=3) over the training image

would be different with a different training image.

1.2.2 Template dimension

Similar to the categorical case, with a 250×250 training image grid and a 100×100 simulation grid, the default search template dimension of 11×11 provides good results (Figure 1.19). The best results are again provided when the dimensions of the search template and the inner patch are in reasonable proportions: respectively (11×11 and 7×7) or (15×15 and 11×11). Decreasing too much the inner patch dimension for a given search template dimension does not improve the realization and is time consuming. Indeed the (11×11 and 3×3) realization of Figure 1.19 is crispier than the (11×11 and 7×7) realization. Thus and similar to the categorical case, a reasonable ratio of Search template vs Inner patch should be kept.
CHAPTER 1. SENSITIVITY ANALYSIS OF FILTERSIM

Figure 1.17: Realizations with K-mean partition (Initialization number increasing from 100 to 1000) Row 2: Histogram of the realizations

Figure 1.18: Time for one realization as function of the Initialization number of the K-mean partition
Figure 1.19: Realizations for different template parameter inputs
1.2.3 Number of multigrids

Sensitivity analysis is now performed on the number of multigrids. On Figure 1.20 with only one multigrid the artifact patch can be seen on the realization, it is less apparent on the realization performed with 3 multigrids. Moreover the simulated channels are less continuous on the single multigrid realization. In the case of a $250 \times 250$ training image used here with a $100 \times 100$ simulation grid, the default number 3 for multigrids appears good. Note that for all these realizations the histogram is well reproduced but the spatial variance is too small (variance=0.44 in the training image). Indeed there is almost no values higher than 3.5 in these four realizations.

![Figure 1.20: Realizations for different number of multigrids with default parameters](image)

1.2.4 Cmin to control the partition process

On Figure 1.21 whatever the value of Cmin is, the histogram of the realization is well reproduced with mean closed to the target 1.91. For all these realizations, the variance is still smaller than the target 0.44. Decreasing Cmin slightly reduces the patch effect: the patch $7 \times 7$ can be clearly seen on the realization with a Cmin=40, it is less apparent on the realizations with Cmin smaller than 10.
1.2. 2D SENSITIVITY ANALYSIS WITH CONTINUOUS VARIABLES

Figure 1.21: Realizations for different number of Cmin with default parameters

1.2.5 Servosystem

On Figure 1.22, the servosystem appears to have no influence. There appear only ergodic fluctuations from one realization to another one. The histogram is roughly reproduced for all realizations but the spatial variance is still too small even for the maximum servosystem=1 where the higher variance=0.41 is only due to a lucky ergodic fluctuation. The reproduction of the histogram and especially its mean and variance need to be improved. The second chapter of this thesis will precisely focus on the reproduction of the target histogram.
1.2.6 Ergodicity

Frozen parameters: Cross partition with Secondary=2, search template=11 × 11, inner patch=7 × 7, 3 multigrids and Cmin=10.

From one realization to another one, there are necessarily ergodic fluctuations due to the simulation field not being (infinitely) large. On Figure 1.23, the point wise average (E-type) of 100 non-conditional realizations using the training image from Figure 1.1 converges to the target mean=1.91 whatever the size of the simulation grid is. On Figure 1.24, the mean of the spatial variance of 20 realizations is 0.32, thus not converging to the training image variance 0.44. On Figure 1.24, when the simulation grid is larger (250 × 250) the ergodic fluctuations of the spatial variance are smaller, as expected, but the mean of that variance (0.31) is still too small compared to that (0.44) of the training image.

Variogram reproduction

- Sill reproduction: As seen on Figure 1.24, the spatial variance of the simulated realizations is most of the time too small compared to that of the training image. On Figure 1.25, increasing the number of parents in the partition process (e.g. increasing
1.2. 2D SENSITIVITY ANALYSIS WITH CONTINUOUS VARIABLES

Figure 1.23: E-type of 100 non conditional realizations for a 100*100 and a 250*250 grid

Figure 1.24: Spatial variance fluctuations over 20 non conditional realizations for a 100*100 and a 250*250 grid
the Initialization number) allows to have a better reproduction of that spatial variance (sill of the variogram). Indeed the realization with a Initialization number of 10 has the higher spatial variance 0.39. On Figure 1.26, the realizations generated for Figure 1.25 are reproducing well the channel continuity and the target mean but with a too small variance.

- Range reproduction: To reproduce the range of the variogram of the training image, the size of the simulation grid has to be larger than four times the value of the range. On Figure 1.26, with a 100 × 100 grid the range of 40 is not reached. It is reached with the larger 250 × 250 grid, but with a smaller variance (lower sill).

Reproducing the range of the variogram and decreasing the ergodic fluctuations calls for a large simulation grid and is time consuming. Reproducing the spatial variance of the training image calls also for increasing the Initialization number but this is time consuming.

Figure 1.25: Variogram for the 250 × 250 training image (in red) and for 3 realizations on a 250 × 250 grid for different Initialization numbers with a cross partition method (Initialization=10 (green), Initialization=6 (blue), Initialization=3 (black)
1.2. **2D SENSITIVITY ANALYSIS WITH CONTINUOUS VARIABLES**

Figure 1.26: Realizations on a $250 \times 250$ for an Initialization number of 3, 6 and 10 with default parameters

Figure 1.27: Variogram of 10 realizations on a $100 \times 100$ and a $250 \times 250$ grid (training image variogram in red)
1.3 Conclusions

Geological patterns and target proportions of the training image are reasonably well reproduced by the filtersim realizations for both categorical and continuous variables. Quality of the realizations and CPU time are very sensitive to the partition method and especially to the Initialization number that should be set first and carefully. The cross partition method is faster than the K-mean partition. Default parameters are proposed but should be adapted to the dimensions of the training image and the grid simulation. A search template of half the size of the simulation grid at the coarser grid scale is a reasonable choice. The reproduction of the target proportions (categorical case) or the histogram (continuous case) is not well controlled by the present servosystem. This problem is handled in the second part of this thesis.
Chapter 2

Reproduction of histogram with filtersim

2.1 Introduction

In the original version of the code, the user enters a specific proportion target and the servosystem parameter aims at reproducing this target. To reproduce that target proportion during the sequential simulation process, instead of randomly picking a pattern from a prototype class, a pattern is selected such that its patching draw the current realization proportion closer to the target [Zhang (2006)]. As discussed previously in the introduction on Figures 2(a) and 2(b), this parameter fails to reproduce the target mean and the variance of the training image. This servosystem suffers from any or all of three major problems:

- the servosystem is not practically nor theoretically correct to ask for a target that is different from that of the training image. The algorithm is designed to reproduce some of the multiple-point statistics of the training image and those include one-point statistics (histogram and especially the mean, the variance and the extremes values).

- it is limited to reproduce only the mean, hence critical extreme values proportions need not be reproduced which would explain the bias on the mean and on the variance seen on Figures 2(a) and 2(b).

- the servosystem presently coded is efficient only if there is enough variability in mean values within a prototype class. At the same time there should not be variability within
To address the first problem, it is recommended to pre-process the training image with an histogram transform before using \textsc{filtersim}. The target histogram then becomes the histogram of the training image. Next, we suggest removing the original servosystem and replace it by an algorithm based on histogram transform so that to reproduce not only the target mean or proportions, but also an entire target histogram so that the variance and the proportion of extremes values can be better matched. An histogram transform is performed on the simulated values before the last finest grid simulation. Such a transformation enables to pass to the last grid simulated values closer to the target histogram. As all the nodes are re-simulated on the last grid, the final realizations will still honor all hard data. On Figures 2.1(a) and 2.1(b) this method (K-mean classification method and an histogram transform) shows a better reproduction of the one-point statistics of the training image compared to Figures 2(a) and 2(b).


2.2 Histogram reproduction

2.2.1 Base case

The training image from Figure 2.2 is used to generate unconditional realizations on a 100 \times 100 grid using filtersim. The reproduction of one-point statistics (mean, variance, maximum value, etc...) is analysed over 100 realizations. The default parameters used in all the next realizations unless otherwise noted are given in Table 2.1.

![Figure 2.2: 2D continuous training image](image)

<table>
<thead>
<tr>
<th>General menu</th>
<th>Advanced menu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Template: 11 11 7</td>
<td>Servosystem Factor: 0 (inactive)</td>
</tr>
<tr>
<td>Inner Patch: 7 7 5</td>
<td>Number of Multigrids: 3</td>
</tr>
<tr>
<td>Target Mean: 1.93</td>
<td>Min # Replicates for Each Multigrid: 10 10 10</td>
</tr>
<tr>
<td></td>
<td>Weights to Hard, Patch and Others: 0.5 0.3 0.2</td>
</tr>
<tr>
<td></td>
<td>Pattern Partition Method: Partition with K-mean</td>
</tr>
<tr>
<td></td>
<td>Maximum Number of Clusters:</td>
</tr>
<tr>
<td></td>
<td>Initialization: 200</td>
</tr>
<tr>
<td></td>
<td>Secondary Partition: 2</td>
</tr>
</tbody>
</table>

Table 2.1: Default parameters

Figures 2.3(a), 2.3(b) and 2.3(c) show the histogram of the 100 simulated means, variances and maximum values when using the previous default parameters. Figure 2.4 shows the three first realizations out of the 100 run.
CHAPTER 2. REPRODUCTION OF HISTOGRAM WITH FILTERSIM

(a) Histogram of mean  
(b) Histogram of variance  
(c) Histogram of maximum

Figure 2.3: Default parameters

(a) First realization  
(b) Second realization  
(c) Third realization

Figure 2.4: Realizations with default parameters
To check that the number of 100 realizations is enough to avoid major ergodic fluctuations, 50 and 200 realizations were performed with the same parameters. In Table 2.2, the mean of the spatial variance over 200 realizations appears reasonably close to that of 100 realizations and differs only slightly from that of 50 realizations.

<table>
<thead>
<tr>
<th>number of realizations</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean of the variance</td>
<td>0.828</td>
<td>0.836</td>
<td>0.835</td>
</tr>
<tr>
<td>mean of the maximum</td>
<td>5.58</td>
<td>5.53</td>
<td>5.55</td>
</tr>
</tbody>
</table>

Table 2.2: Sensitivity of spatial variance and maximum value to number of realizations

Some ergodic fluctuations are expected between different realizations using the same input (same parameters and same data), and these fluctuations can be used to evaluate uncertainty of global statistics such as spatial mean and variance. However, these ergodic fluctuations should not lead to biases. In Table 2.6, for 100 realizations the mean of the 100 simulated means is 1.80 different from the target 1.93 and the mean of the 100 simulated variances is 0.84 significantly lower than the target 1.06. In Table 2.7, the standard deviation, characterizing ergodic fluctuations among the 100 realizations, for the spatial mean and the variance are respectively 0.14 and 0.13 which is relatively high.

2.2.2 The servosystem and its problems

As shown in Figure 2(b), the spatial variance of anyone of the 100 realizations is too low and the maximum values are also too low. The bias in the variance is thus likely to come from the lack of extreme simulated values in the realizations. This is true for the three realizations of Figure 2.4 where there are fewer red high-valued locations than in the training image of Figure 2.2. It may be critical to reproduce such extreme values that could either be barriers or channels impacting flow simulations. On Figure 2.5, the histogram of the means of the training prototypes reveals that both classification methods fail to regroup patterns with extremes values in a single or a small number of classes. The problem is worst for the cross partition method where the prototype with the highest mean is only 4.05 even though there are 1709 classes. For the k-mean partition method with only 200 classes, the prototype with the highest mean is 5.42. This latter higher value explains why the k-mean partition method gives better results than the cross partition (see summary Table 2.6). This problem is to be expected because the classification filters are not designed to capture extreme values.
especially on the coarse grid. Consequently during the simulation process, it is difficult to choose a specific pattern with extreme values. The pattern selection also depends on the distance used. Using the Manhattan distance does not help to draw a pattern with extreme values; an Euclidean distance with squared differences would do better.

![Histogram of mean of prototypes for cross partition](image1)

![Histogram of mean of prototypes for K-mean partition](image2)

(a) Histogram of mean of prototypes for cross partition  (b) Histogram of mean of prototypes for K-mean partition

Figure 2.5: Classification at fine grid

The inability of the classification to regroup training patterns with extreme values within a small number of classes is more important at the coarsest scale (first grid being simulated), because of the large scale templates that cannot detect local variations. At this scale, templates and filters can only capture the large scale structure of the training image. Even with the k-mean partition method, the prototype with the highest mean is only 3.04 on Figure 2.6 compare to the training maximum value 7.15.

![Histogram of mean of prototypes for K-mean partition at the first grid](image3)

Figure 2.6: Histogram of mean of prototypes for K-mean partition at the first grid
2.2. HISTOGRAM REPRODUCTION

2.2.3 Histogram transform method

Using the histogram of the training image from Figure 1(b) as target, the values simulated at the next to last grid are transformed. For example, if there are 3 multiple grids, the histogram transform is applied after the second grid has been simulated. Then the last grid simulation is performed as usual with all the nodes being re-simulated. Figure 2.7 shows the steps of this process for one specific realization. Figure 2.7(d) shows the final realization corresponding to Figure 2.8(a).

(a) Coarse grid
(b) Medium grid
(c) Medium grid after histogram transformation
(d) Fine grid

Figure 2.7: Different steps of the simulation process

On Figures 2.8(a), 2.8(b) and 2.8(c), there appears more high values (red locations) than in the base case on Figures 2.4(a), 2.4(b) and 2.4(c). On Figure 2.4(b), the realization displays the same shape than that on Figure 2.4(c) due to the rank preserving character of the histogram transform. On Figure 2.9, the final histograms of the realizations are seen to be close to that of the training image on Figure 1(b). In Table 2.6, the reproduction of the mean, the variance and high values is seen to be much better than when using the default parameters.
CHAPTER 2. REPRODUCTION OF HISTOGRAM WITH FILTERSIM

(a) First realization  
(b) Second realization  
(c) Third realization

Figure 2.8: Realizations with histogram transformation

(a) First realization  
(b) Second realization  
(c) Third realization

Figure 2.9: Histogram of realizations with histogram transformation
2.2. HISTOGRAM REPRODUCTION

### 2.2.4 Miscellaneous approaches

The reproduction of the histogram can be further improved especially by combining the histogram transform with a better distance calculation method. Indeed in Table 2.6, the reproduction of the mean and the variance, although improved, is not perfect due to the simulation at the last grid.

**Score based distance**

Filter score values are used instead of pixel values in the calculation of the distance between a DEV and a class prototype (Wu, 2006). In Table 2.3, the CPU time required for 100 realizations is about the same when using score distance, since most of the simulation time is taken by the preliminary steps of classification, see Table 2.4.

On Figures 2.10(a), 2.10(b) and 2.10(c) realizations look similar to that obtained with default parameters and the pixel distance. In Table 2.6, the reproduction of the mean, the variance and high values is slightly better than when using the pixel distance.

<table>
<thead>
<tr>
<th></th>
<th>1 realization</th>
<th>100 realizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixels distance</td>
<td>529s</td>
<td>593s</td>
</tr>
<tr>
<td>Score distance</td>
<td>524s</td>
<td>549s</td>
</tr>
</tbody>
</table>

Table 2.3: Time for classification + realizations

<table>
<thead>
<tr>
<th>Grid</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarse</td>
<td>75s</td>
</tr>
<tr>
<td>Medium</td>
<td>100s</td>
</tr>
<tr>
<td>Fine</td>
<td>345s</td>
</tr>
<tr>
<td>Total</td>
<td>520s</td>
</tr>
</tbody>
</table>

Table 2.4: Time for classification at different grids

**Histogram transform and score based distance**

The histogram transform method is combined with the score distance method. At the next to last grid, the dual template concept (Arpat, 2004) is used to fill the non-informed locations. Then using the target histogram of the training image of Figure 1(b), the data values of the fully informed dual template grid are transformed. Finally using score distance, simulation on the last grid is performed with these transformed data values. On Figure 2.13, the dual template grid after histogram transformation show more high values. Applying the histogram
CHAPTER 2. REPRODUCTION OF HISTOGRAM WITH FILTERSIM

(a) First realization  (b) Second realization  (c) Third realization

Figure 2.10: Realizations with score distance transform on the dual grid has more impact than applying it just on the penultimate grid because the dual grid has twice more pixel values. In Table 2.6, the reproduction of the mean, the variance and the maximum value is seen to be better than when using the pixel based distance.

(a) First realization  (b) Second realization  (c) Third realization

Figure 2.11: Realizations with histogram transformation + distance score

Euclidean distance

Instead of using the Manhattan distance from equation 2.1, the Euclidean distance equation 2.2 is used to calculate the distance between a DEV and a prototype. By weighting less the impact of extremes values in comparison to the Euclidean distance, the Manhattan distance is less effective in selecting prototypes with such extremes values especially when there are many data found in the DEV on the last grid scale. In Table 2.6, realizations with Euclidean distance show a better reproduction of the mean, the variance and of the maximum value.

\[ d(x,y) = \sum_{i=1}^{n} |x_i - y_i| \]  

(2.1)
2.2. HISTOGRAM REPRODUCTION

Figure 2.12: Histogram of realizations with histogram transformation + distance score

(a) First realization   (b) Second realization   (c) Third realization

Figure 2.13: Different steps of the simulation process

(a) Coarse grid   (b) Medium grid
(c) Dual template   (d) Dual template after histogram transform
CHAPTER 2. REPRODUCTION OF HISTOGRAM WITH FILTERSIM

\[ d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]  

(2.2)

**Histogram transform + Euclidean distance**

In Table 2.6, realizations performed with both the histogram transform and the Euclidean distance show a better reproduction of the mean, the variance and the maximum values than the base case. However compared to the histogram transform only, using in addition the Euclidean distance does not improve much further the histogram reproduction.

**Larger realization size**

The simulation grid is now set to the same size as the training image: 250 × 250. On Table 2.7 it is seen that by increasing the size of the simulation grid ergodic fluctuations decrease. However the bias in the mean and variance still remains because it is essentially due to the classification process.

![Realizations on 250 × 250](image)

(a) First realization  
(b) Second realization

Figure 2.14: Realizations on 250 × 250

**Larger training image**

A larger training image provides more replicates of the training patterns. The X and Y axes are used as axes of symmetry to get mirror symmetry of the original TI from Figure 1(a). On
2.2. HISTOGRAM REPRODUCTION

Figure 2.15: 250 x 250 simulation grid

(a) Histogram of mean
(b) Histogram of variance

Figure 2.16: Mirror TI (500 x 500)

This new image is used as a training image to run 100 realizations on a 100 x 100 simulation grid. From a pattern variety point of view, it is equivalent to get all the original patterns + the original patterns rotated by 90, 180 and 270 + a few negligible patterns from the junction between images. In Table 2.6, the mean and the variance are equally poorly reproduced. Indeed as the training image is larger, the classification of patterns with extreme values is even more difficult. On Figure 2.18, the histograms of 3 realizations do not reproduce the target histogram on Figure 1(a).
CHAPTER 2. REPRODUCTION OF HISTOGRAM WITH FILTERSIM

<table>
<thead>
<tr>
<th>Coarse</th>
<th>Medium</th>
<th>Fine</th>
<th>Total classification</th>
<th>Classification + 1 realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1143s</td>
<td>1132s</td>
<td>2156s</td>
<td>4431s</td>
<td>4598s</td>
</tr>
</tbody>
</table>

Table 2.5: Time for classification and realization

(a) First realization  
(b) Second realization  
(c) Third realization

Figure 2.17: Realizations with larger TI

(a) First realization  
(b) Second realization  
(c) Third realization

Figure 2.18: Histogram of realizations with larger TI
### 2.2. HISTOGRAM REPRODUCTION

#### 2.2.5 Evaluation of multiple-point statistics realizations

**One-point statistics**

All the statistics of the realizations performed are gathered in Table 2.6. It shows that the one-point statistics are better reproduced with the histogram transform. The mean, the variance and the maximum value are better reproduced with the histogram transform. These targets are also better reproduced with the Euclidean distance and the score distance. The best reproduction is done by pooling histogram transform, Euclidean distance and score distance. In Table 2.7, ergodic fluctuations are shown to be lower with the histogram transform method: this is as expected.

<table>
<thead>
<tr>
<th></th>
<th>mean of the mean</th>
<th>mean of the variance</th>
<th>mean of the max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target from the TI</strong></td>
<td>1.93</td>
<td>1.06</td>
<td>7.15</td>
</tr>
<tr>
<td><strong>Default parameters</strong></td>
<td>1.809</td>
<td>0.836</td>
<td>5.525</td>
</tr>
<tr>
<td>Cross partition</td>
<td>1.820</td>
<td>0.630</td>
<td>4.984</td>
</tr>
<tr>
<td>Histogram transform</td>
<td>1.864</td>
<td>0.912</td>
<td>5.715</td>
</tr>
<tr>
<td>Distance score</td>
<td>1.834</td>
<td>0.874</td>
<td>5.725</td>
</tr>
<tr>
<td>Histogram transform</td>
<td>1.892</td>
<td>0.946</td>
<td>6.054</td>
</tr>
<tr>
<td>+ Distance score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>1.920</td>
<td>0.868</td>
<td>5.898</td>
</tr>
<tr>
<td>Histogram transform</td>
<td>1.894</td>
<td>0.895</td>
<td>5.870</td>
</tr>
<tr>
<td>+ Euclidean distance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larger training image</td>
<td>1.631</td>
<td>0.837</td>
<td>5.229</td>
</tr>
<tr>
<td>Larger grid</td>
<td>1.791</td>
<td>0.815</td>
<td>6.406</td>
</tr>
</tbody>
</table>

Table 2.6: Reproduction of basic statistics
<table>
<thead>
<tr>
<th>interfere with the std of the mean</th>
<th>std of the variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default parameters</td>
<td>0.139</td>
</tr>
<tr>
<td>Cross partition</td>
<td>0.101</td>
</tr>
<tr>
<td>Histogram transform</td>
<td>0.043</td>
</tr>
<tr>
<td>Distance score</td>
<td>0.131</td>
</tr>
<tr>
<td>Histogram transform + Distance score</td>
<td>0.038</td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>0.148</td>
</tr>
<tr>
<td>Histogram transform + Euclidean distance</td>
<td>0.044</td>
</tr>
<tr>
<td>Histogram transform + Distance score + Euclidean distance</td>
<td>0.040</td>
</tr>
<tr>
<td>Larger training image</td>
<td>0.175</td>
</tr>
<tr>
<td>Larger grid</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Table 2.7: Measure of ergodic fluctuations

Two-point statistics

Two-points statistics (variograms) of realizations are compared to those of the training image through variogram maps. The variogram map depicts the spatial variability in all the directions $h$ with at his center $\gamma(0) = 0$:

$$\gamma(h) = \frac{1}{2} E \left( |Z(u) - Z(u+h)|^2 \right)$$

(2.3)

Variogram maps are calculated using the Fast Fourier Transform, Marcotte (1996). On Figure 2.19, the variogram map of the training image indicates a higher NE-SW connectivity and some NN-SE periodic structure, essentially due to the channels. In the following, variogram maps are limited to a lag of 50 to ensure that there are enough pair of data in a 100 × 100 realization. To compare these 2-points variogram maps, the effect of the variance is removed by normalizing each variogram value by the means of the respective variogram maps. Figure 2.20(b) shows the average variogram map of 100 standardized variogram maps corresponding to 100 realizations run with the default parameters. It is very close to the variogram map of the training image given on Figure 2.20(a). On Figure 2.20(c) and 2.20(d),
the variogram map of realizations using histogram transform + score distance are similar to that of realizations using cross partition. Thus 2-points statistics are reasonably well reproduced in all these realizations.

![TI variogram map (lag of 125)](image)

Figure 2.19: TI variogram map (lag of 125)

**Effective permeability**

The horizontal effective permeability of the training image is compared to those of 100 realizations. Indeed engineers are more interested in flow properties of the model than reproduction of variogram models. The global horizontal effective permeability $k_e$ is calculated for each realization as an horizontal harmonic average over the $N X k_{i,j}$ permeability values followed by a vertical arithmetic average over the $N Z k_{i,j}$ permeability values:

$$k_e = \frac{1}{N Z} \sum_{i=1}^{N Z} \frac{N X}{\sum_{j=1}^{N X} k_{i,j}}$$

The effective permeability of the training image is 1.385. On Table 2.8, the effective permeability of realizations using the default parameters is too low and ergodic fluctuations are relatively important. By using histogram transformation, reproduction of the training image effective permeability is improved and the ergodic fluctuations are reduced.
CHAPTER 2. REPRODUCTION OF HISTOGRAM WITH FILTERSIM

Figure 2.20: Variogram maps for 100 realizations (lag of 50)

<table>
<thead>
<tr>
<th></th>
<th>mean of the permeability</th>
<th>std of the permeability</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>1.385</td>
<td></td>
</tr>
<tr>
<td>Default parameters</td>
<td>1.357</td>
<td>0.144</td>
</tr>
<tr>
<td>Distance score</td>
<td>1.370</td>
<td>0.138</td>
</tr>
<tr>
<td>Histogram transform</td>
<td>1.382</td>
<td>0.057</td>
</tr>
<tr>
<td>Histogram transform + score</td>
<td>1.407</td>
<td>0.058</td>
</tr>
<tr>
<td>Cross partition</td>
<td>1.458</td>
<td>0.114</td>
</tr>
</tbody>
</table>

Table 2.8: Measure of horizontal effective permeability
2.3 Conclusion and future work

Filtersim simulated realizations fail to reproduce the histogram of the training image. The cross partition method is worse at such reproduction than the K-mean partition method. Filtersim simulated realizations using the proposed histogram transform reproduce better the histogram of the training image and especially its extreme values. This reproduction can further be improved by using an Euclidean distance and the score-based distance to find the closest class prototype to any specific conditioning data event. This histogram transform does not affect the reproduction of two-point statistics.
Chapter 3

Impact of mps on estimation

Filtersim generates multiple equally probable simulated realizations that aim at reproducing the geological patterns of the input training image. Once the filtersim algorithm is properly calibrated, these realizations can be used to assess local uncertainty. That assessment is compared to that provided by direct kriging.

3.1 Two facies simulation

3.1.1 Data analysis

The reference image of dimensions 150 × 150 pixels given in Figure 3.1(a) depicts EW channels elongated in a mud background (mud=74%, sand=26%). A representative set of 50 conditioning hard data is sampled and shown in Figure 3.1(b) (mud=72%, sand=28%). Figure 3.1(c) gives the reference image variograms along the EW and NS directions; these are fit by an anisotropic spherical model with zero nugget effect with EW range=28 and NS range=8.

Geological interpretation could have provided the large training image of Figure 3.2(a) with similar EW elongated channels and proportions (mud=72% sand=28%) tuned to be close to the sample proportions. On Figure 3.2(b) the training image variogram is fitted with a two-structures model: zero nugget effect, a first isotropic exponential structure of range 16, and a second anisotropic spherical structure with ranges 34 and 8. This variogram model is similar to that adopted for the reference image. Only statistics that would be available in practice are used hereafter, as extracted from the training image (proportions...
3.1. TWO FACIES SIMULATION

and variogram model). All simulations and estimations are conditioned to the same set of 50 sample data taken from the reference image Figure 3.1(b).

3.1.2 Kriging estimate

The training image channel proportion (sand=0.28%) and variogram model are used for simple indicator kriging. After some sensitivity analysis, a data search ellipsoid of 50*50 is retained with a maximum number of conditioning data=20. On Figure 3.3(a) the simple kriging estimate map appears smooth with high and low estimated values located around the corresponding hard data. At other locations further away from the data, that is at locations with high kriging variance on Figure 3.3(b), the kriging estimate is close to the input mean (0.28). On Figure 3.3(c), the histogram of the kriging estimate is no more binary though it is unbiased (mean=0.28), it shows a large reduction in variance compared to that of the training image: $0.02 << 0.20 = 0.28 \times (1 - 0.28)$. From the sole kriging map of Figure 3.3(a), one may not be able to identify existence of continuous channels.

3.1.3 E-type of sisim realizations

Using the training image sand indicator variogram and the 50 sample data, we generated 30 conditional realizations using Sequential Indicator SIMulation (program sisim in Deutsch and Journel (1998)). Figure 3.4(a) shows one typical realization: it is seen to fail reproducing the channel continuity yet it honors exactly the 50 hard data. By averaging 30 such SISIM realizations, we get the E-type estimated map of Figure 3.4(b). The same exercise is repeated with 200 realizations resulting in the E-type of Figure 3.4(c). As these simulations are all conditioned to the same hard data, both E-type maps also honor exactly these hard data. As expected these E-type maps converge to the kriging map of Figure 3.3(a) as the number of realizations increases. Similar to the kriging estimate, we fail to see the channel continuity, even though the input variogram model does reflect the higher EW continuity. The histograms of the E-type maps also fail to reproduce the binary character of both the reference and training images, see Figure 3.4(d) and 3.4(e).
CHAPTER 3. IMPACT OF MPS ON ESTIMATION

(a) Reference image 150*150
(b) Sample of 50 hard data
(c) Variogram of the reference in the EW and NS directions

Figure 3.1: Reference data set
3.1. TWO FACIES SIMULATION

(a) Training image (250*250)

(b) Variogram of the training image in the EW and NS directions

Figure 3.2: Training image
CHAPTER 3. IMPACT OF MPS ON ESTIMATION

Figure 3.3: Kriging

(a) Kriging estimate map
(b) Kriging variance map
(c) Kriging estimate histogram
3.1. TWO FACIES SIMULATION

(a) Sisim realization

(b) E-type on 30 realizations

(c) E-type on 200 realizations

(d) E-type histogram (30 realizations)

(e) E-type histogram (200 realizations)

Figure 3.4: Sisim
3.1.4 E-type of filtersim realizations

Using the default set of parameters of program filtersim, we ran 30, then 200 realizations conditioned to the same 50 hard data. Each of the resulting realizations is now structurally similar to the training image and has proportion close to the target (28% sand). Figure 3.5(a) gives one typical realization that resembles the reference image except for some channel discontinuity. That specific realization has sand proportion equal to 26%. On Figure 3.5(b) and 3.5(c), the two E-type maps provide indication of presence of continuous EW channels as opposed to the kriging map of Figure 3.3(a). The E-type means, respectively 0.26 and 0.27, are close to the target training image value 0.28. As expected for any average-type map, there is still a large reduction of the spatial variance that is a smoothing effect, although not as severe as for kriging or the sisim-based E-type estimates of Figure 3.4(c).

3.1.5 Estimates comparison: local accuracy

To compare the reference binary variable $I$ with the simple kriging estimate $P_\star^K$, the sisim E-type estimate $P_\star^S$ and the filtersim E-type estimate $P_\star^F$, these estimates are transformed into binary indicator maps $I_\star^K$, $I_\star^S$ and $I_\star^F$ using threshold values $z_K$, $z_S$ and $z_F$ defined such as to filter out the small residual bias:

\[ P_\star^K \rightarrow I_\star^K = 1 \quad if \quad Z_\star^K > z_K \quad with \quad z_K \quad s.t \quad \frac{1}{N} \sum_{i=1}^{n} I_\star^K(i) = \frac{1}{N} \sum_{i=1}^{n} I(i) \quad (3.1) \]

\[ P_\star^S \rightarrow I_\star^S = 1 \quad if \quad Z_\star^S > z_S \quad with \quad z_S \quad s.t \quad \frac{1}{N} \sum_{i=1}^{n} I_\star^S(i) = \frac{1}{N} \sum_{i=1}^{n} I(i) \quad (3.2) \]

\[ P_\star^F \rightarrow I_\star^F = 1 \quad if \quad Z_\star^F > z_F \quad with \quad z_F \quad s.t \quad \frac{1}{N} \sum_{i=1}^{n} I_\star^F(i) = \frac{1}{N} \sum_{i=1}^{n} I(i) \quad (3.3) \]

where $N = 150 \times 150 = 22,500$ is the field size, $I(i) = 1$ if the reference pixel at location $i$ is sand. On figure 3.1(a), the reference has 26% of sand, thus: $\frac{1}{N} \sum_{i=1}^{n} I(i) = 0.26$. Comparing figures 3.6(a)(b) and (c), the indicator map of the filtersim E-type displays a better EW continuity than the kriging or the sisim E-type indicator maps.

Next the local accuracy of these three estimates is compared using the estimation variance:

\[ \sigma^2 = \frac{1}{N} \sum_{i=1}^{n} | I_\star(i) - I(i) |^2 \quad (3.4) \]
3.1. TWO FACIES SIMULATION

(a) filtersim realization

(b) E-type on 30 realizations

(c) E-type on 200 realizations

(d) E-type histogram (30 realizations)

(e) E-type histogram (200 realizations)

Figure 3.5: Filtersim
resulting in:

\[
\sigma_S^2 = 0.28 > \sigma_K^2 = 0.27 > \sigma_F^2 = 0.25
\]

which indicates that the filtersim E-type estimate is slightly more locally accurate than the kriging estimate and the sisim estimate E-type.

Another measure of local accuracy is the mean absolute error: \(|I_K^* - I|\), \(|I_S^* - I|\) and \(|I_F^* - I|\). For the filtersim E-type estimate, it is 0.30 a value significantly smaller than that for the kriging estimate (0.34) and for the sisim E-type estimate (0.35).

### 3.1.6 Estimates comparison: connectivity

A multiple-point statistics such as the \(N\)-order connectivity allows comparing the presence and the continuity of channels in the different estimated images. The \(N\)-order rectilinear connectivity function \(\phi_N\) measures the frequency of \(N\) continuous connected pixels along a straight line.

\[
\phi_N = E\left\{\prod_{i=1}^{N} I(u+i)\right\}
\]  

(3.5)

We use the bandwidth connectivity program developed by Krishnan that accounts for curvilinearity of local structures by allowing an angle tolerance, [Krishnan and Journel (2003)](#). The connectivity statistics of each image is standardized to filter out differences in the image means. This connectivity is calculated first on the 2-facies simulated images generated by sisim on Figure [3.4(a)] and by filtersim on Figure [3.5(a)]. On Figure [3.7(a)] the strict rectilinear EW connectivity is calculated with zero direction tolerance, i.e with zero bandwidth. There appears no significant difference between the filtersim realization (red curve), the sisim realization (blue curve) and the reference model (black curve); this is because the strictly rectilinear connectivity measure fails in all three cases to account for the channel curvilinearity. By increasing the bandwidth on Figure [3.7(b)] and [3.7(c)] the channel curvilinearity is accounted for; now the filtersim realization displays higher curvilinear connectivity than the sisim realization, although lesser than that of the reference image.

To compare the connectivity of the two E-type images and the kriging estimated map, we must consider binary indicator maps obtained by thresholding these estimated images. Using the histograms of Figure [3.3(c)] [3.4(d)] and [3.5(d)] we select the single probability threshold value 0.4 that reveals the high probability values for presence of sand. It is not necessary to ensure the same proportions in the resulting three binary images as was the case for the
Figure 3.6: Indicator maps ensuring the same global proportion of sand
local accuracy comparison, because the connectivity measure used is standardized. Figure 3.8 gives the resulting binary images after thresholding. These three binary images honor exactly the same thresholded hard data. The sand bodies revealed by these binary images all fail to display the continuity of the reference channels: they are less sinuous, less continuous and thicker.

Next, we use the previous connectivity program with a bandwidth=15 to compare the binary images of Figure 3.8(b), 3.8(d) and 3.8(f). On Figure 3.9(a) when using only 30 realizations, the filtersim E-type shows already a better curvilinear connectivity than the sisim E-type (blue curve) or the kriging estimate (green). On Figure 3.9(b) when using 200 realizations, the connectivity of the filtersim E-type (red) is now significantly higher than the connectivity of the sisim E-type (blue), or kriging (green), although it remains lower than that of the reference model (black). However, the channel bodies provided by the truncated E-type in Figure 3.8(f) do not reflect the reference channel continuity. Recall that reproduction of continuity is the goal of a simulated realization not of an E-type or a kriging map. On Figure 3.10, the EW variogram of the filtersim E-type present a more marked hole effect than the EW variograms of the kriging image or the sisim E-type; this hole effect reflects the sinuosity of the channels.
3.1. **TWO FACIES SIMULATION**

Figure 3.7: Connectivity measure for one sisim realization (blue), one filtersim realization (red) and the reference image (dark)
Figure 3.8: Binary images after thresholding at 0.4 the kriging estimate, the sisim and filtersim E-type maps
3.1. TWO FACIES SIMULATION

(a) EW bandwidth connectivity (30 realizations for each E-type)

(b) EW bandwidth connectivity (200 realizations for each E-type)

Figure 3.9: EW bandwidth connectivity (bandwidth=15)
(a) Variogram of the kriging estimate map

(b) Variogram of the sisim E-type (30 realizations)
3.1. TWO FACIES SIMULATION

(c) Variogram of the filtersim E-type (30 realizations)

Figure 3.10: Variograms in EW and NS directions

3.1.7 Binary case conclusions

Only the filtersim realizations (Figure 3.5(a)) were able to reproduce reasonably the continuity of the training image channels. The local accuracy of the filtersim E-type is slightly better than that of the kriging estimate. As for E-type average maps, only the filtersim E-type (Figure 3.5(b) and 3.5(c)) gives indication of presence of channels; no such indication is detected on the sisim E-type or kriging estimate maps. The bandwidth connectivity of the filtersim E-type map is significantly higher than that of the sisim E-type map and closer to the reference connectivity. However and as expected, the channels on the filtersim E-type map still fail to reflect the crisp geometry of the channels present on the reference model. This is due to the smoothing effect of all least-square type estimates, E-type maps are no exception.
3.2 Four facies simulation

In this section, we repeat the previous study, but considering a four facies case.

3.2.1 Data analysis

Figure 3.11(a) shows the 2D reference image with four facies: shale=58% (dark blue=1), channel sand=17% (light-blue=2), levee=17% (yellow=3), crevasse (brown=4)=8%. On Figure 3.11(c), the selected training image displays similar geological features with proportions: shale=45%, channel sand=20%, levee=20% and crevasse=15%. For indicator kriging and simulation, we retrieve and model the indicator variogram of each facies from the training image. Figure 3.11(b) gives the set of 50 hard data retained with sample proportions: shale=58%, channel sand=16%, levee=12% and crevasse=14%, see also the summary table in Figure 3.11(d). These sample proportions are close to the reference proportions and are used as target for estimation and simulation.

3.2.2 Kriging estimate

Using the four indicator variograms modeled from the training image and the sample target proportions, simple indicator kriging was performed for each of the four facies resulting in Figures 3.12(a), (b), (c) and (d). These maps all honor exactly the corresponding 50 indicator hard data. At any location, the sum of the four facies estimated probabilities is equal to one. On Figure 3.13 the histograms of the four facies indicator estimates show a mean close to the target proportion for each facies: for facies 1, mean=0.54 (target=0.58); for facies 2, mean=0.18 (target=0.16); for facies 3, mean=0.13 (target=0.12); for facies 4, mean=0.15 (target=0.14). From these histograms we can define thresholds to transform the continuous indicator kriging maps into binary maps identifying exactly the target proportions, then compute their connectivity. Figure 3.14 shows the corresponding binary maps.

3.2.3 Sisim realization

Figure 3.15(a) gives one typical realization of sisim using the four variograms modeled from the training image and using for target the sample data proportions. This realization is conditioned to the 50 sample data of Figure 3.11(b). The geological features of the reference
image (sand channels with attached levees) are not reproduced. The proportions, however, are reasonably well reproduced, see table of Figure 3.15(c).

### 3.2.4 Filtersim realization

Figure 3.15(b) gives one typical realization of filtersim using for target the same sample data proportions and conditioned to the same 50 sample data. The geological features of the training image are now reasonably well reproduced. Figures 3.16(a), (b), (c) and (d) give the four facies E-type maps built from 200 filtersim realizations. These E-type maps were turned into binary maps by applying threshold values ensuring reproduction of the target facies proportions; the resulting binary maps are given in Figure 3.17.

### 3.2.5 Estimates comparison: local accuracy

To compare, for each facies $k$, the local accuracy of the kriging estimate $P_{K,k}$ and the filtersim E-type estimate $P_{F,k}$ to the reference value $I_k$, we utilized the binary indicator maps as defined in equations (1), (2), (3) and shown in Figure 3.17. For each facies, the estimation variances between the estimated indicator maps and the reference map are calculated with equation (4), resulting in:

- **Sand (facies 1):** $\sigma_{K,1}^2 = 0.18 > \sigma_{F,1}^2 = 0.10$
- **Shale (facies 2):** $\sigma_{K,2}^2 = 0.42 > \sigma_{F,2}^2 = 0.37$
- **Levee (facies 3):** $\sigma_{K,3}^2 = 0.73 > \sigma_{F,3}^2 = 0.65$
- **Crevasse (facies 4):** $\sigma_{K,4}^2 = 0.43 < \sigma_{F,4}^2 = 0.52$

After weighting by the target proportions $t_k$, the average estimation variance is:

$$\sigma_k^2 = \frac{1}{N} \sum_{k=1}^{4} t_k \sigma_{K,k}^2 = 0.32 > \sigma_F^2 = \frac{1}{N} \sum_{k=1}^{4} t_k \sigma_{F,k}^2 = 0.27$$  (3.6)

Another measure of local accuracy is the absolute error difference $|I_{K,k} - I_k|$ between the two binary indicator maps $I_{K,k}$ and $I_{F,k}$ and the reference map $I_k$ for each facies see Table 3.2.5. By comparing for each facies the means of these maps, we see that for facies 1, 2 and...
Chapter 3. Impact of MPS on Estimation

3, the local accuracy of the filtersim E-type is better.

In summary the filtersim E-type appears more locally accurate than the kriging estimate.

<table>
<thead>
<tr>
<th></th>
<th>Shale</th>
<th>Sand</th>
<th>Levee</th>
<th>Crevasse</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>I_{F,k} - I_k</td>
<td>$</td>
<td>0.28</td>
<td>0.58</td>
</tr>
<tr>
<td>$</td>
<td>I_{K,k} - I_k</td>
<td>$</td>
<td>0.34</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 3.1: Measure of local accuracy

3.2.6 Estimates comparison: Connectivity

We compute the curvilinear connectivity with a bandwidth=15 for each of the binary maps obtained by thresholding the indicator kriging map (Figure 3.14) and the filtersim E-type map (Figure 3.17). These connectivity results are compared to the connectivity of the reference map. On Figure 3.18 for shale (facies 1) and sand (facies 2) the filtersim E-type (red curves) yields a higher EW continuity very close to that of the reference connectivity. For levee (facies 3) and crevasse (facies 4), both the filtersim E-type and the kriging estimates fail to reproduce the reference connectivity.

As in the binary case study (part 1), these results confirm the better continuity reproduction of the filtersim algorithm, especially for facies that are characterized by a high connectivity.

3.3 Conclusion

In these two examples, the filtersim simulated realizations are shown to reproduce reasonably well the short scale geological patterns depicted by the training image and more poorly the large scale structures. The local accuracy of the filtersim E-type is better than that of the indicator kriging estimate particularly for facies with large scale continuity (channels). The bandwidth connectivity statistics confirms the advantage of the filtersim E-type over the indicator kriging; this is significant only for facies with high connectivity.
3.3. CONCLUSION

(a) Reference image 150*150
(b) Sample of 50 hard data
(c) Training image
(d) Facies proportions

Figure 3.11: Data set
CHAPTER 3. IMPACT OF MPS ON ESTIMATION

(a) Kriging map for shale (facies 1)

(b) Kriging map for sand (facies 2)

(c) Kriging map for levee (facies 3)

(d) Kriging map for crevasse (facies 4)

Figure 3.12: Indicator kriging maps
3.3. CONCLUSION

(a) Indicator kriging of shale, mean = 0.54

(b) Indicator kriging of sand, mean = 0.18

(c) Indicator kriging of levee, mean = 0.13

(d) Indicator kriging of crevasse, mean = 0.15

Figure 3.13: Histograms of indicator kriging maps
Chapter 3. Impact of MPS on Estimation

(a) Shale (facies 1), mean=0.58
(b) Sand (facies 2), mean=0.16
(c) Levee (facies 3), mean=0.12
(d) Crevasse (facies 4), mean=0.14

Figure 3.14: Binary images after thresholding the indicator kriging maps
3.3. CONCLUSION

(a) SISIM realization

(b) filtersim realization

<table>
<thead>
<tr>
<th>Facies</th>
<th>SISIM</th>
<th>Filtersim</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.58</td>
<td>0.45</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.17</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>0.08</td>
<td>0.15</td>
</tr>
</tbody>
</table>

(c) SISIM and filtersim realizations proportions

Figure 3.15: Sisim and filtersim realizations
Figure 3.16: Filtersim E-type on 200 realizations
3.3. CONCLUSION

(a) Shale (facies 1) 
(b) Sand (facies 2) 
(c) Levee (facies 3) 
(d) Crevasse (facies 4) 

Figure 3.17: Binary images after thresholding filters in E-type maps, ensuring target proportions
Figure 3.18: EW connectivity for the reference (black), the filtersim E-type (red), the kriging estimate (green) and the TI (blue) for each facies
Bibliography


