Exploring Direct Sampling and Iterative Spatial Resampling in History Matching

Matz Haugen, Grergoire Mariethoz and Tapan Mukerji

Department of Energy Resources Engineering
Stanford University

Abstract

We explore multiple methods to history match a porosity field by using both production data and a secondary data set. In this paper, seismic impedance is used as secondary data but any data set with identical dimensions as the primary data can be used. The task is formulated as an inverse problem where the production data is estimated through an upscaled flow simulator. Co-located permeability values are obtained though an empirical relationship from the well logs. The secondary seismic data is treated in two ways. Firstly, it is used as conditioning data to further constrain the porosity simulation. This means that the production data is the only data that has a likelihood function. In the second method, porosity and seismic impedance are simulated simultaneously. This corresponds to a likelihood function with combined production and seismic data. If the second method is used, a more constrained porosity field is obtained in which channels follow the seismic more closely, whereas when the first method is used the channels are slightly thinner and more spread out. Because increased amount of conditioning data yields a more constrained *a posteriori* distribution it is likely that the optimization takes requires less steps to reach an optimum compared with the case where both seismic and production data is matched.

1 Introduction

In the past, a significant effort has been made to match reservoir models to production data like water-cut and cumulative oil production Caers (2007); Hoffman (2005),Caers and Hoffman (2006). Methods have also been proposed to combine primary data (e.g. production rates) with secondary data Mosegaard and Tarantola (1995) in order to further constrain the inverse problem and mitigate ill-posedness. In this paper, the reservoir model is defined as a spatial field with porosity, permeability and seismic impedance.

In the framework of inverse problem theory, one wants to find the spatial distribution of the attributes of interest that corresponds to certain models

outputs (typically measured quantities). The relationship between the input parameters distribution and the models outputs is given by a possibly non-linear and complex forward model. The forward model can either be based on physical laws or on a statistical relationship between the input parameters and the model outputs. For example, computing pressure based on permeability data and given boundary conditions using Darcy's law is a physically-based forward problem, and deriving permeability from porosity using the Hagen-Poiseuille relationship is a statistical forward problem.

Either of the methods can be faster than the other depending on the type of problem. However, once one method is chosen, the degree of accuracy (or geologic realism) of the model can sometimes still be adjusted. Thus, it often becomes a trade-off between the level of accuracy and the time it takes before a satisfactory set of parameters have been chosen. The complexity of the problem also plays a role in this choice. Large models take longer to simulate.

If a statistical model is to be chosen, the geologic consistency depends on the technique that is used. One possible geologic scenario is a subsurface channel system. This type of geology is sometimes non-stationary and in order to make simulations that reproduce the channel-like features, multiple-point geostatistical algorithms are required. For this, several methods are available Strebelle (2002); Zhang et al. (2006a); Mariethoz et al. (2010b). These methods all require a training-image. Usually, increased geological consistency comes at the price of computational cost. For example in the case of Strebelle (2002), RAM and CPU demand are directly proportional to the template size, and with the DS algorithm, CPU cost is related to the acceptance criterion concerning the misfit to the neighbor locations (see Section 4). Therefore a trade-off regarding simulation time and geologic realism comes into play.

When considering the actual simulations, there are a couple of ways one can approach the fact that both production and seismic data are known observations. If a multivariate training image is available, it can provide an experimental statistical relationship between the variables. If one of the variables is known, this statistical relationship can be used to guide the simulation of the other. This can be the case if seismic data is known everywhere and a bivariate training image is available to provide the statistical relationship with unknown porosity. This way of incorporating secondary data uses it as conditioning data. Instead of creating new simulations of both primary and secondary data sets, the secondary variable is assumed to be known and can be used to constrain the primary dataset using the known relationship between variables. This happens during the actual geostatistical simulation. Therefore, the likelihood function has a single term, $L_p(m)$, considering only the production data. In some cases, this could be an advantage because two terms in one objective function can yield competing effects. An extreme case would be that one variable reaches a local minimum while the other is in a local maximum. The training image would be unchanged, still containing both variables, but the realization only outputs one simulated variable.

A different approach would be where, instead of using the seismic data as conditioning data, it is also simulated and matched to the observed seismic.

Consequentially, two likelihood functions could be incorporated into an objective function and used in an optimization problem, one for production data and one for seismic data. Call the respective production and seismic likelihood functions $L_p(m)$ and $L_s(m)$.

In this paper, the terms included are cumulative water cut, cumulative oil production and seismic impedance, where the first two are bundled into $L_p(m)$. The last term ($L_s(m)$) is excluded when using seismic as conditioning data with a statistical relationship between both variables.

The remainder of the paper is structured by first giving an explanation of the methodology used in Section 2. The specifics of the sampling of the posterior in Baysian terms (or optimization) is discussed in Section 3 followed by a description of the geospatial algorithm used (Section 4). A discussion of the reference field is given in Section 5.2 and Section 6 shows the obtained results using both methods discussed above regarding the secondary data. Lastly, a conclusion is presented in Section 7.

2 Methodology

The overall workflow of the optimization problem is shown in Figure 1. An initial first guess is perturbed and the resulting new model is evaluated. Two tasks that are performed simultaneously for each forward problem evaluation. Only the reservoir simulation is relevant when the seismic is used as conditioning data. Once the objective value is calculated, a rejection rule defined in Equation 2 is applied. If accepted, the model is again perturbed and evaluated in a next iteration. It is up to the user to decide a stopping criterion for the procedure. In this particular case, a fixed number of 100 iterations was set. Other stopping criteria could also be applied, e.g. stop the process if no likelihood has been found that is better than the current after some number of iterations.

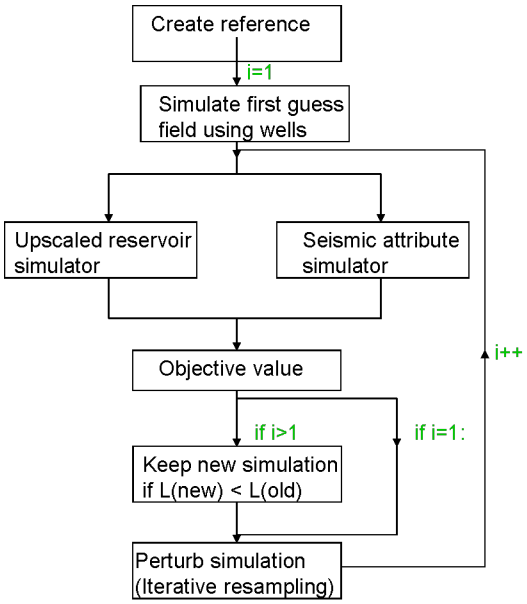

Figure 1: Graphical representation of the workflow. i is the current number of iterations. In terms of a Markov Process, i represents the i^{th} step.

3 History Matching

In order to best sample the distribution of possible solutions given the observations, stochastic methods are used in an iterative process that eventually reaches a local minimum by matching the production history. Each starting point is a sample from the prior distribution and the subsequent iterations are gradient-free modifications to the current model. Because this is an under-constrained problem, there are can be many possible local minima and effort should be made to find as many as possible in order to quantify uncertainty. In this paper, a method called Iterative Spatial Resampling (ISR) will be used to perform the spatial perturbation, a method initially proposed by Mariethoz et al. (2010a).

The principle of ISR is as follows. In a given iteration, a fraction of the simulated variable is retained and used as conditioning data in the subsequent realization. The larger the fraction of data retained, the more similar the next iteration will be to the previous (parent) realization. A graphical representation of the iterative process is shown in Figure 2.

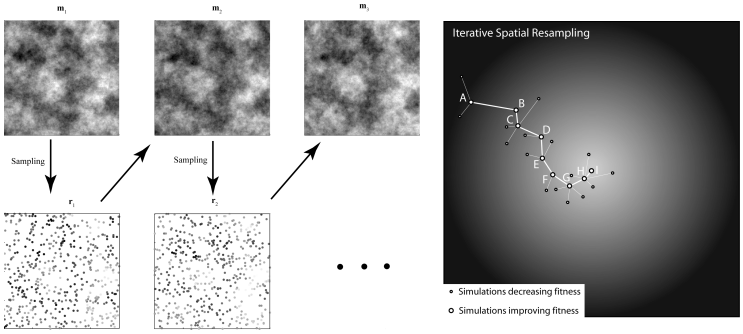

Figure 2: A graphical representation of Iterative Spatial Resampling. Left: A normally distributed Gaussian field undergoing iterative resampling. Right: The value of the objective function projected onto two dimensions showing a convergence to a local minimum.

In a Bayesian framework, we want to find the posterior likelihood distribution

$$f(m|d) = \frac{f(d|m)f(m)}{f(d)} \tag{1}$$

where d is the data and m is the model, $f(m)$ is the prior distribution without any data and $f(d|m)$ is the likelihood function that usually involves a forward problem. It defines a probabilistic relationship between the observed state variables d (the data) and the model m Mariethoz et al. (2010a). We define an objective function (or likelihood function) $L(m)$ that shows the misfit between the given model and the reference. In this paper, the perturbation will follow the following rule when going from one model to the next,

$$\begin{array}{ll} \text{if } L(m^*) > L(m_i), & m_{i+1} = m^* \\ \text{otherwise,} & m_{i+1} = m_i \end{array} \tag{2}$$

where m^* is the new model with perturbed values of porosity and m_i is the current model. It is important to note that is does not yield the theoretically accurate posterior distribution, but it is an approximation. The full rejection rules that are consistent with the Baysian framework is as follows,

$$\begin{array}{ll} \text{if } L(m^*) > L(m_i), & \text{move from } m_{i+1} = m^* \\ \text{if } L(m^*) < L(m_i), & \text{accept } m^* \text{ with probability } L(m^*)/L(m_i) \end{array} \tag{3}$$

which is proven by Mosegaard and Tarantola (1995).

4 Direct Sampling

Direct sampling is a multiple-point geostatistical simulation technique introduced by Mariethoz et al. (2010b). With the traditional MPS approaches (e.g.

Strebelle (2002); Zhang et al. (2006b); Straubhaar et al. (2008)), the training image is usually scanned and all pixels configurations of a certain size (the template size) are stored in a catalogue of data events having a tree or a list structure. This structure is then used to compute the conditional probabilities at each simulated node. Since the memory load increases dramatically with the size of the template and the number of facies, the use of MPS is generally restricted to relatively small problems with limited structural complexity. Moreover, the template is often not large enough to capture large-scale structures, and multi-grids have been introduced to palliate this problem by simulating the large-scale structures first, and later the small-scale features. Limited memory usage restricts the approach to the simulation of univariate fields.

The DS algorithm takes a different approach. Instead of counting and storing the configurations found in the training image, it directly samples the training image for a given data event. The method is based on a sampling method introduced by Shannon (1948), strictly equivalent to the original MPS algorithm of Guardiano and Srivastava (1993), but that does not explicitly need to compute conditional probabilities and therefore does not need to store them. Starting from an initial randomly located point, the training image is scanned. For each of the successive samples, the distance between the data event observed in the simulation and the one sampled from the training image is calculated. If the distance is lower than a given threshold, the sampling process is stopped and the value at the central node of the data event in the training image is directly used for the simulation. Since nothing is stored, neighborhoods can have virtually any size and are not restricted to a template, making the use of multiple-grids unnecessary. In fact, multiple-grids are replaced by a continuous variation in the size of data events during the simulation process. This reduces artifacts and leads to a simpler implementation.

Similarly to other methods, DS can reproduce the structures of complex training images and deal with a wide range of non-stationary problems. Additionally, it can simulate both discrete and continuous variables by choosing specific distances between data events, and can deal with specific cases of non-stationarity. But the richest feature of DS is that multivariate data configurations can be considered, allowing to generate realizations presenting a given multivariate multiple-point dependence. The ability to treat multiple variables simultaneously in a multiple-point framework allows co-simulation without formulating the dependency between the simulated attributes. Instead, the multiple-point dependence is reproduced as it is in the multivariate training image.

The multivariate possibility is perhaps the most important reason why the DS algorithm is used in this study. With earlier algorithms, this could be more cumbersome to implement, e.g. using FILTERSIM Zhang et al. (2006a). If the option of multivariate simulation were not there, the secondary variable would have to be estimated with a forward deterministic geophysical simulation and porosity and impedance would be simulated separately, while the DS algorithm simulates both variables simultaneously.

One important input parameter for DS is the distance characterizing the

mismatch between a data event in the realization and another one in the training image. Here we use a Euclidian distance normalized to a range of $[0, 1]$, with a distance threshold of 0.06. . This means that the first point found in the training image that has a lower distance than 0.06 is placed at the simulated location in the realization. In other words, data events presenting a normalized mismatch of 6% or less are considered acceptable. The outcome of the realizations is quite dependent on this threshold value in terms of quality of the channels (e.g. connectivity) and also has a significant effect on the time required for each simulation. It is naturally in our interest to have as small a simulation time as possible and still get reasonable, geology-preserving results. Choosing a lower threshold increases the computational burden. A low threshold forces exact reproduction of the trining image's patterns, which may not be desirable (for example if we know that the training image is not perfectly representative of the field to simulate) On the other hand, if the threshold is too high, geological structures will deviate from the realistic scenario.

5 Case Study

The properties that are simulated are porosity and seismic impedance. From porosity, an empirical relationship with permeability can be established. It allows obtaining permeability on the whole field, assuming that the porosity-permeability relationship is accurate. The relationship deduced is shown below in Equation 4 which is a relationship obtained from the Stanford VI well logsEcheverra et al. (2009),

$$K = a * \phi^3/(1-\phi)^2 \tag{4}$$

where $a \approx 4.35 \cdot 10^4$; and ϕ is porosity and K is permeability.

Both the reference field and the training image are taken from the 3D synthetic field Stanford VI. The geology is fluvial and the porosity field is shown below. The correlation coefficient between impedance and porosity is about 0.5 for the training image, in line with the reference. Both the reference and the training images are taken from slices with the correlation coefficient quite high in value (upper quartile), and the reason that a number of slices have significantly lower values is because the impedance is highly influenced by the adjacent slices, i.e. the frequency spectrum of the seismic is relatively high. This mean that the weight of the seismic should take this fact into account.

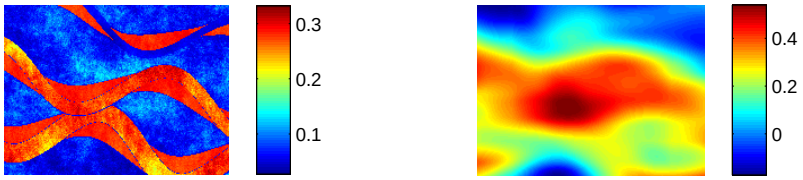

Figure 3: Training image with porosity on the left and seismic impedance on the right

The task of choosing a training image is challenging. Because the reference in a real situation is not known, the parameters characterizing the field must be determined from geological background knowledge. In the case studied here, parameters like channel width and sinuosity should be emphasized. The actual location of the channels with respect to the wells locations should not be relevant as this is handled by the DS algorithm when including the conditional data points.

5.1 Simulations conditioned to well data only

With 20 realizations using the training image in Figure 3, the resulting four images are shown in Figure 4 where one realization is only conditioned to the well data (left) and the other also conditions with the seismic impedance (right). Furthermore, the means of the two configurations are the top images while the variance is shown in the bottom two. The images that are only constrained to the well data (left) have a significantly larger variance in the areas not near the wells compared with the well areas. This variance is decreased when a secondary variable is used to further constrain the simulation. In fact, the secondary variable seems to constrain the simulation so much that the output seems to be only one channel cutting across the area that in the reference has two channels. The remaining two well areas are reduced to islands of high porosity, which are consistent with the training image that also presents such islands.

One could argue that the training image is too small and not enough variation is included. This could be remedied by potentially combining several layers from the 3D Stanford VI data set to get a larger 2D training image which could supply the algorithm with enough variation in the channel shapes.

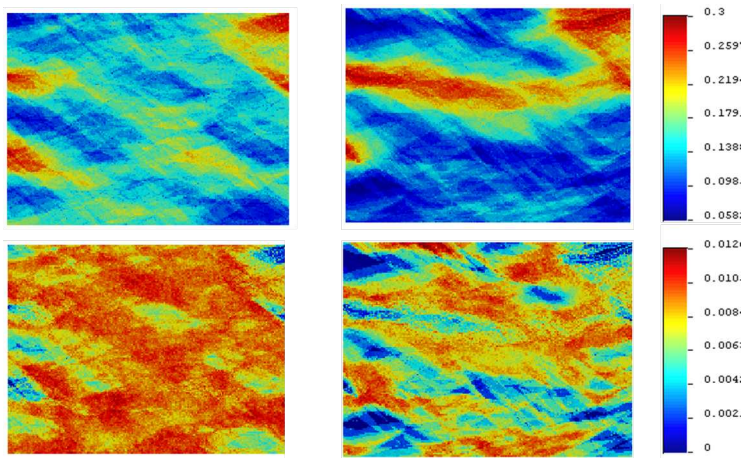

Figure 4: Top: Mean of 20 simulations conditioned to well data only on the left and conditioned to both well data and seismic impedance on the right. Bottom: Associated variances of the top fields. The training image that was used is shown in Figure 3.

5.2 The Reference

For the methodological purpose of this paper, a two dimensional field will be used instead of a three dimensional field to reduce the time of each iteration. By reducing the spatial dimensionality, only seismic impedance will be used as secondary data. One could envision a 3D case where cross-well or diffraction tomography could also be incorporated into the objective function. Other seismic attributes could replace the impedance, but because the impedance does not constrain the reservoir too much, it is a convenient secondary variable.

The real porosity and seismic impedance are depicted in Figure 5. In this case, the correlation coefficient between the two is about 0.5. It is important to note that if the co-located correlation coefficient between the two variables is too high it might over-constrain the porosity simulation and render the whole optimization process useless after the first guess because the first guess would already be quite good.

As can be seen in Figure 5), it is not at all obvious that there are two channels when only looking at the impedance. This is significant for the DS algorithm because the information given in the training image is now ambiguous since the frequency spectrum of the secondary variable does not capture the finer scaled features. Consequently, the algorithm cannot distinguish between one larger and two smaller channels. In other words, the dimensions of the channel are such that they fall within a Fresnel zone and are below the lateral resolution.

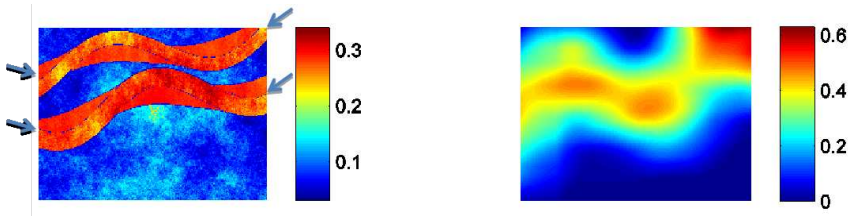

Figure 5: Reference porosity and seismic impedance. The arrows indicate well locations where porosity and permeability values are known.

In order to decrease the simulation time, upscaling is performed on the model. In particular, the 150 by 200 pixels grid is upscaled to a 30 by 40 pixels grid. The water saturation profiles at different times are depicted below in Figure 6. The porosity is upscaled using an arithmetic mean while the permeability is calculated using a pressure solver written by Deutsch (1987). [1] For the actual simulation, the Stanford General Purpose Simulator (GPRS) is used, but because of the black-box approach concerning the simulator, any simulator could be used.

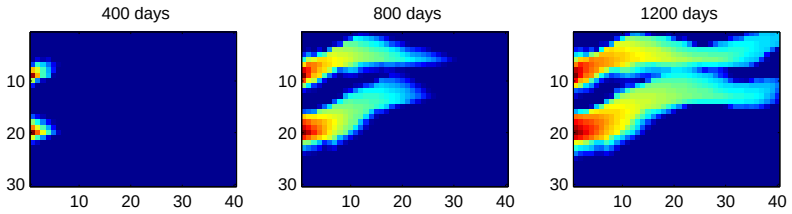

Figure 6: Saturation profiles using a coarse model.

The fine scale version is depicted in Figure 7, showing considerably more detail in the water saturations. However, even though there is significant visual difference, the actual production results do not differ greatly. A juxtaposition of the two cases is shown in Figure 8. In terms of simulation times, the fine scaled model takes about 25 minutes to run while the coarse scaled takes 7 seconds.

[1] The upscaler is called FlowSim and is available on the SCRF website as Fortran code. It applies a constant pressure gradient in one direction with no boundary conditions on the other faces of each of the blocks in the model.

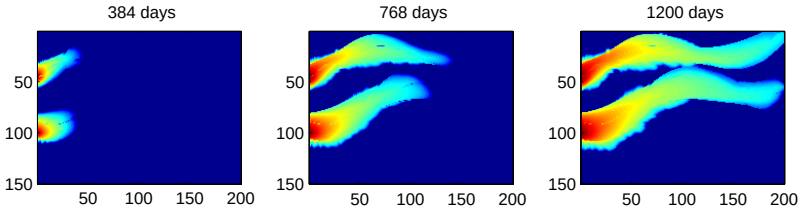

Figure 7: Saturation profiles using a fine scaled model.

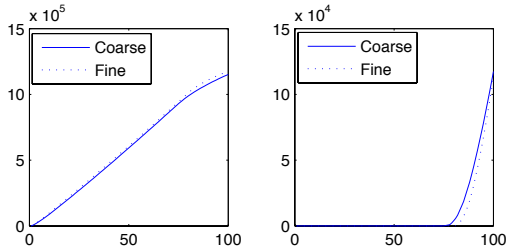

Figure 8: cumulative oil and water production

6 Results

6.1 Simulating both seismic and porosity

After 100 ISR iterations, the objective function reaches a local minimum. Consider first the seismic as a simulated secondary variable in the geostatistical simulation (thus not conditional). The results are presented by first showing the prior simulations in Figure 9 before starting ISR whereas Figure 10 shows the final result. The top two images depict a randomly chosen simulation of variable 1 (left) and 2 (right), the middle two show the mean of the 10 realizations and the bottom two show the associated variances of the two variables. We can see that the final results (Fig. 10) shows a more connected channel profile and a slightly lower variance in the channel area. As mentioned in Section 5.2, the distance threshold for the misfit between the simulated point and the point of interest in the training image is 0.06.

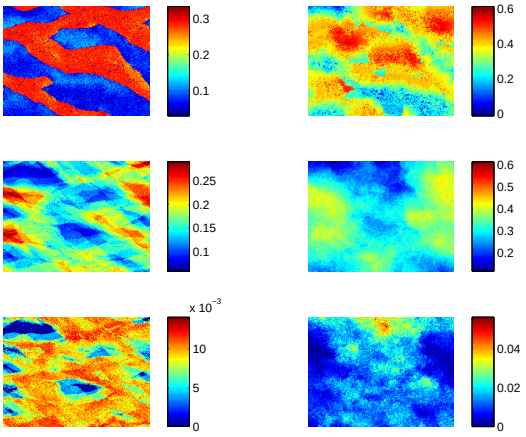

Figure 9: Initial results before ISR has been implemented. The right figures concern the primary variable, porosity, while the left figures concern the secondary variable (seismic impedance). Top: one realization of porosity (left) and the corresponding seismic impedance (right). Middle: The mean of 10 realizations. Bottom: Variance of the 10 realizations.

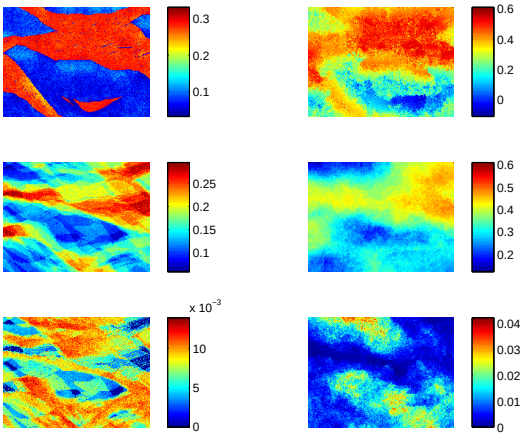

Figure 10: Final results after ISR has been implemented. The right figures concern the primary variable, porosity, while the left figures concern the secondary variable (seismic impedance). Top: one realization of porosity (left) and the corresponding seismic impedance (right). Middle: The mean of 10 realizations. Bottom: Variance of the 10 realizations.

The corresponding cumulative production values are shown in the following figure (Fig. 11). As can be seen, there is a convergence toward the reference in

all three figures.

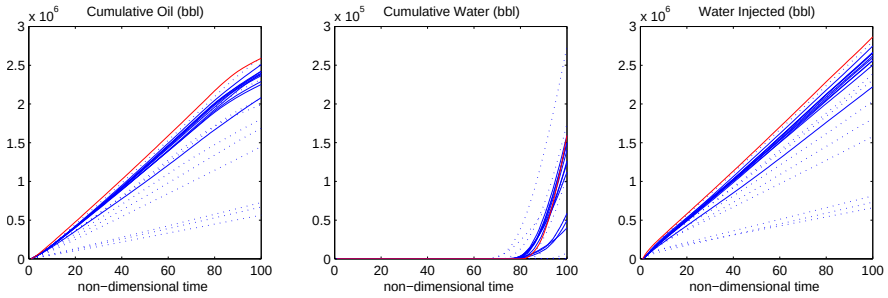

Figure 11: The production/injection rates of oil/water (left/middle) with the water injection levels on the right. The reference is in red, the solid line is the final iteration and the dotted lines are the first guess simulations.

To show that there are many possible scenarios that can produce similar production data a set of water saturation images are shown below taken at various times during the appraisal (Fig. 12). We see here that there are two veins that merge into a larger one, as opposed to two entirely separated channels. With the given information the two possible scenarios are difficult to discern.

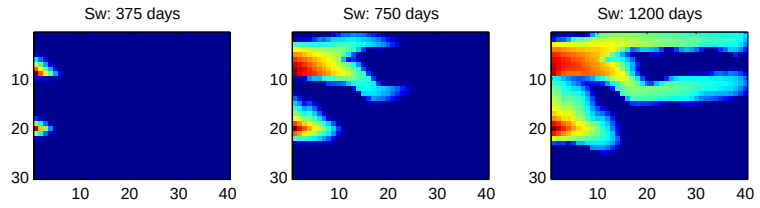

Figure 12: Water saturation profiles one randomly chosen model after ISR has been implemented and simulating both porosity and permeability.

6.2 Simulating porosity only

If the seismic is used as conditional data the results show slightly different results. In Figure 13, two realizations are shown on top with the mean and variance of 10 realizations are shown on the bottom left and right. The mean shows a channel profile that has the majority of the channel higher upward compared to Figure 10. This is the result of a tighter constrained prior distribution when including the secondary variable as conditional to the primary.

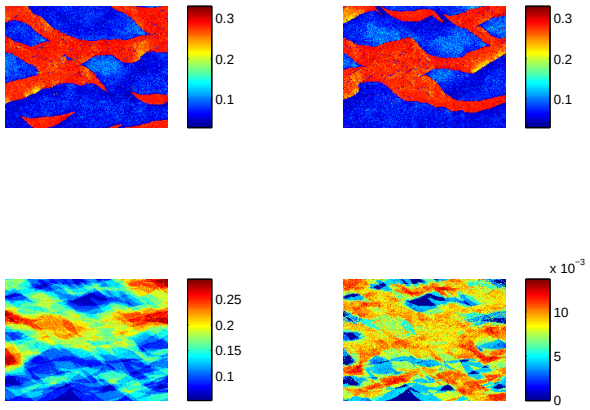

Figure 13: Top: Two realizations using the secondary variable as conditioning data. Bottom: The mean (left) and the variance (right) associated with 10 realizations of the prior

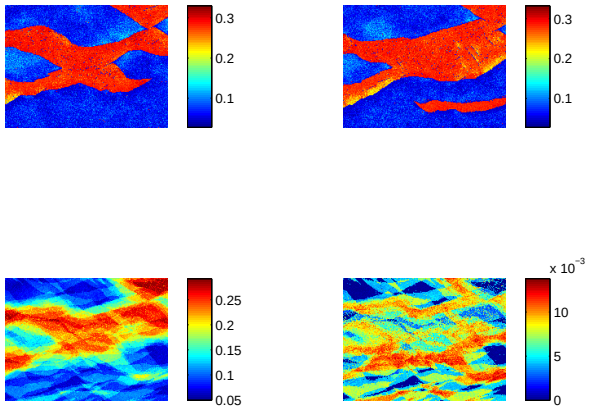

Figure 14: Top: Two realizations using the secondary variable as conditioning data after 100 iterations of ISR. Bottom: The mean (left) and the variance (right) associated with 10 realizations of the approximated posterior

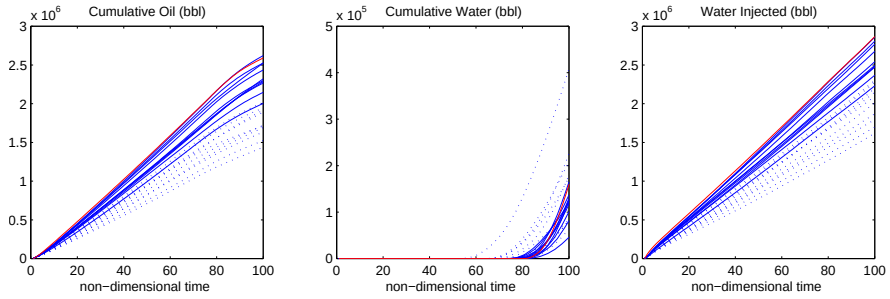

Figure 15: The production/injection rates of oil/water (left/middle) with the water injection levels on the right. The reference is in red, the solid line is the final iteration and the dotted lines are the first guess simulations.

For completeness the water saturation profiles are shown in the following figure (Fig. 16).

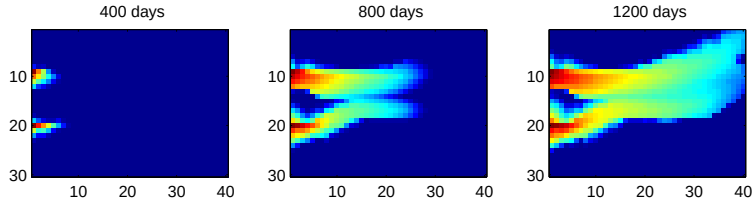

Figure 16: Water saturation profiles one one randomly chosen model after ISR has been implemented using seismic as conditioning data.

If the cutoff (see Section 4) is reduced to 0.05, the cumulative production values obtained are shown in Figure 17 using seismic data as conditioning data. This does not show a significant improvement from the previous result and verifies the hypothesis presented in Section 4.

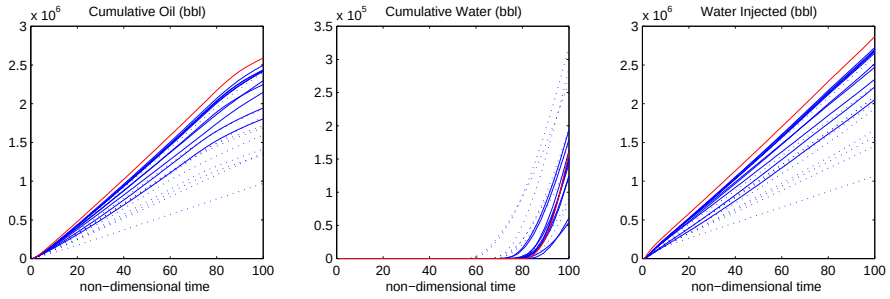

Figure 17: The production/injection rates of oil/water (left/middle) with the water injection levels on the right using a cutoff of 0.05 instead of 0.06. The reference is in red, the solid line is the final iteration and the dotted lines are the first guess simulations.

7 Conclusion

As geostatistical algorithms like the DS algorithm opens up the possibility of simulating multiple variables simultaneously, the history matching process can be more constrained. Therefore, in certain cases complex and expensive physical forward problems can be replaced by statistical forward problems.

For most algorithms, there are many parameters that can be changed and part of the challenge is to find the ones that give the best result. In the DS algorithm, these are parameters the size of the search neighborhood, the weight of each variable and the distance threshold. When using the type of seismic impedance data that has been seen in this paper, the relative weight of the secondary variable could be set as inferior to the primary. The quantification of the weights should be based on the amount of new information that is obtained when introducing a new variable and could be a challenging task from a heuristic stand point. More elaborate methods to quantify these weights could involve the Tau/Nu model Krishnan et al. (2008).

Furthermore, because there is a high wavelength cutoff in the seismic impedance data's frequency spectrum, the individual channels are hard to map out when looking at only the secondary variable. It seems as if the secondary variable shows only one connected channel due to the long wavelength. Further exploration will define what effect that the different input parameters have on this artifact.

References

Caers, J., 2007. Comparing the gradual deformation with the probability perturbation method for solving inverse problems. Mathematical Geology 39 (1), 27–52.

Caers, J., Hoffman, T., 2006. The probability perturbation method: A new look at bayesian inverse modeling. Mathematical Geology 38 (1), 27–52.

Deutsch, 1987. Flowsim.
URL http://www.ualberta.ca/~cdeutsch/

Echeverra, D., Mukerji, T., Santos, E., 2009. Robust scheme for inversion of seismic and production data for reservoir facies modeling. SEG Houston 2009 International Exposition and Annual Meeting, 2432–2436.

Guardiano, F., Srivastava, M., 1993. Multivariate geostatistics: Beyond bivariate moments, Geostatistics-Troia. Kluwier Academic, pp. 133–144.

Hoffman, T., 2005. Geologically consistent history matching of facies models. Ph.D. thesis, Stanford University.

Krishnan, S., Boucher, A., Journel, A. G., 2008. Evaluating information redundancy through the tau model. Quantitative Geology and Geostatistics.

Mariethoz, G., Renard, P., Caers, J., 2010a. Bayesian inverse problem and optimization with iterative spatial resampling. Submitted to Water Resources Research.

Mariethoz, G., Renard, P., Straubhaar, J., jan 2010b. The direct sampling method to perform multiple-points geostatistical simulations, submitted to Water Resources Research.

Mosegaard, K., Tarantola, 1995. Monte carlo sampling of solutions to inverse problems. Journal of geophysical Research 100 (B7), 12,431.

Shannon, C. E., 1948. A mathematical theory of communication. The Bell system technical journal 27, 379–423.

Straubhaar, J., Walgenwitz, A., Renard, P., Froidevaux, R., 1-5 Dec. 2008. Optimization issues in 3d multipoint statistics simulation. In: Geostats 2008.

Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology 34 (1), 1–22.

Zhang, T., Switzer, P., Journel, A., 2006a. Filter-based classification of training image patterns for spatial simulation. Mathematical Geology 38 (1), 63.

Zhang, T., Switzer, P., Journel, A., 2006b. Filter-based classification of training image patterns for spatial simulation. Mathematical Geology 38 (1), 63.