Application of Experimental Design in Uncertainty Quantification

Rahman M. Hassanpour

Geostatistical modeling is widely used for spatial modeling in presence of sparse data. There is uncertainty in the input geostatistical parameters that affect the uncertainties in the final models. Experimental design has been used in many areas of engineering such as reservoir engineering. In this short note, the application of experimental design in a geostatistical framework is reviewed.

1. Introduction

Evaluating different recovery processes in a petroleum reservoir requires quantitative geological models that describe all of the features contributing to flow. Geostatistical reservoir modeling is used to quantify geological heterogeneity and to provide a geological model for reservoir simulation studies.

A complete reservoir study requires assessing the uncertainty of the reservoir model. There are many engineering and geological factors involved in the reservoir models that impact recovery predictions. To capture the uncertainty in these factors, multiple geological realizations and scenarios are useful. However, examining the reservoir behavior with numerical simulation for all cases is not viable because of long flow simulation computation time and limited computer resources. In this situation, experimental design techniques can be used first to identify the most important factors that have most significant impact on the response variable and then propose a set of scenarios and experiments to adequately span the space of uncertainty created by the required input parameters. The first purpose of design of experiment is usually a sensitivity analysis followed by uncertainty analysis.

Design of Experiments (DOE) has been used in many areas including reservoir engineering applications. Performance prediction (Chu, 1990; Prada and Cunha, 2008), uncertainty assessment (Damsleth at al., 1992; Bu and Damsleth, 1996; Parada et al., 2008), sensitivity analysis (Willis and White, 2000; Kjonsvik et al., 1994; paper 127 of this report), upscaling (Narayanan et al., 1999) and history matching (Eide et al., 1994; Amudo et al., 2008) are examples of the application of DOE in reservoir engineering. In this paper we discuss DOE in a geostatistical uncertainty assessment framework.

2. Design of Experiment

In general, experimental design theory explains how to sample, over the operation region, the number of cases and levels of input factors used in the simulation work to achieve the most information with the lowest computational costs.

In the experimental design, "factors" are related to the variables that control the result of the experiment (such as inputs to sequential Gaussian simulation) and "levels" are the values of those factors. The "main effect" of a factor refers to the change in the response (such as the model of porosity) produced by a change in the level of that factor. When there are multiple factors involved in an experiment, it is highly possible that the main effect of one factor overlaps with the effect of interactions between other factors. This is called "confounding effects".

Levels are usually linearly scaled to range between -1 to 1. Sometimes a linear scaling is not possible because the corresponding factor may not vary linearly. Permeability is an example of factors that should not be linearly scaled. In this situation a quadratic scaling should be considered (Willis and White, 2000). For the purpose of experimental design study, a list of all factors and levels should be prepared.

The design matrix is a set of factor-value combinations to be processed. There are different methods to generate the design matrix. The vary-one-at-a-time approach involves changing one input variable at a time and compares the resultant change in the response to the base case. This method is inefficient for studies with a large number of factors.

Two and three level factorial design is the most common experimental design technique in industrial applications. In a two-level factorial design, each factor is assigned to its maximum or minimum values in all possible combinations with other factors. For a case with k factors, 2^k experiments must be run. The main application of two-levels factorial design is in screening experiments or sensitivity analysis.

The objective of the screening experiment is to identify those factors that have the most effect on the response variable. This is usually performed in the early stages of a project to remove less important factors. In the three-level factorial design, the minimum (-1), centerpoint (0), and maximum (+1) values are assigned to the factor and 3^k experiments are required for k factors.

A factorial design can be either full or fractional factorial. As the number of factors increases, the number of runs required for full factorial design in both two and three level designs may become intractable. The fractional factorial method can be used to sample some of the most important experimental runs of the full factorial design. In this technique running only a fraction of the full factorial experiment can provide the necessary information about the main effects. For example, in a two-level factorial design with six factors requires 64 runs, only 6 runs correspond to the main factors and only 15 runs correspond to the second-order interactions (combination of 2 out of 6). If assuming that the high order interactions are negligible then only the main and the second order interactions may be obtained with 21 runs (15+6). The interaction orders that need to be neglected are defined by the design resolution. A resolution III design is a design in which no main effects are confounded with any other main effect but main effects are confounded with two-factor interactions and two-factor interactions may be confounded with each other. Resolution IV and V designs can also be considered. Generally, with higher the resolution, less restrictive assumptions are made regarding which interactions are negligible (Montgomery, 2001).

The fractional factorial designs are usually represented in the form of $2^{(k-p)}$ in two-level factorial, where $1/2^p$ represents the fraction of the full two-level factorial design. Similarly, for three-levels the fractional factorial design is represented in the form of $3^{(k-p)}$. For example, $2^{(6-2)}$ is a $1/4th$ fraction of a 2^6 full factorial and $3^{(6-2)}$ is a $1/9th$ fraction of a 3^6 full factorial design.

Half factorial design is explained here in detail from Montgomery (2001) and Antony (2003). Assume a two-levels design with three factors A, B, and C. Table 1 shows the design matrix with all the main, second-order (AB,AC,BC), and third-order (ABC) interactions. In this design 8 combinations are required, but assume that running all combination is not feasible and only half of them can be afforded.

Table 1: Matrix for 2^3 Factorial Design.

Run	A	B	C	AB	AC	BC	ABC
1	1	-1	-1	-1	-1	1	1
2	-1	1	-1	-1	1	-1	1
3	-1	-1	1	1	-1	-1	1
4	1	1	1	1	1	1	1
5	1	1	-1	1	-1	-1	-1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	-1	-1	-1	1	1	1	-1

This design is a 2^{3-1} design. There are $\binom{8}{4} = 70$ options to select 4 runs out of 8 possible runs. A defining relation is needed to select the most important set. In general, the defining relation for a fractional factorial is the set of all columns that are equal to the identity vector. Here, we assume the defining relation is $ABC = I$ which means all runs in Table 1 that their last column is equal to +1 (upper part in the table) should be selected. It can be proven that for the half factorial design, the result would be the same if the defining relation is $ABC = -I$ .The word ABC is called the design generator.

In general, a $2^{(k-p)}$ design can be generated with a 2^k full factorial design and p generators. These generators must be reasonably selected such that the design has the highest possible resolution. Montgomery (2001) provided a complete list of best possible generators that are needed for a $2^{(k-p)}$ fractional factorial design for different factors, runs and resolutions.

Leuangthong (2005) presented a stochastic methodology to optimize the design matrix based on a pre-defined objective function without knowing the response function. This method can be used to calculate the sensitivity terms in a sensitivity study.

Experimental design analysis is usually followed by response surface modeling (RSM). The goal of RSM is to replace the original process by a faster function that will be used to approximate the variation of the computer intensive response variable. A RSM is an empirical fit of experimental or computed responses. In most of the cases, the form of the response function is unknown so the first step in RSM is to find an appropriate form that approximates the true relationship function. Usually a low order polynomial fit with linear regression is utilized. In petroleum applications, quadratic models have worked well in most cases.

3. Applications in Geostatistics

Geostatistical simulation has been widely used in reservoir characterization. The main goal of geostatistical simulation is to assess the uncertainty of the geological models by producing multiple equi-probable realizations. Both local and global uncertainty can be important depending on the future use of the models. Local uncertainty is related to the uncertainty of the property quality at each location in the model and global uncertainty is related to the uncertainty of a global specification of the whole model (such as cumulative oil production given a particular well configuration).

Uncertainty in the geostatistical models is highly related to the input parameters involved in the modeling process. In general, there are two kinds of factors involved in a typical geostatistical modeling workflow; geological factors, and geostatistical modeling factors. Some example of geological and geostatistical factors are tabulated in Table 2.

Geological factors are generally extracted from the delineation wells or conceptual models and are related to the geology of the area. There is typically large uncertainty in geological factors because the sampled data usually account for a very small volume of the modeling area. Geostatistical factors refer to the parameters that are used for geostatistical simulation or estimation. There can also be a large degree of uncertainty in these factors because they are usually inferred from the sparsely sampled data mentioned previously.

Table 2: Examples of uncertain parameters in geostatistical modeling.

Geological Factors	Geostatistical Modeling Factors
Average Porosity	Variogram Range
Average Horizontal Permeability	Variogram Anisotropy
Average Vertical Permeability	Correlation Coefficients
Average Water Saturation	Grid Definition
Facies Proportions	Global Property Distributions
Stratigraphic Surfaces	

Generating multiple geostatistical realizations provides an assessment of model uncertainty (Deutsch, 2002). Ranking methods can be used to select a subset of the realizations to reduce the number of models required to quantify the uncertainty in the geostatistical models. In this method a statistical measure of performance for each realization is calculated. The ranking measure is selected based on the computationally intensive response process. For example, for a steam assisted gravity drainage (SAGD) process, connected pore volume may be used as the ranking measure as it has been shown to be related to oil production. Ranking provides a realistic measure of the space of uncertainty assuming there is no uncertainty in the factors tabulated in Table 2. However, multiple scenarios should be considered when there is uncertainty in the geostatistical modeling parameters.

Considering that there may be multiple factors such as several stratigraphic units, multiple modeling variables and their associated variograms and correlation coefficients, the total number of factors involved in a typical geostatistical modeling could be large, (100+).

Incorporating all parameters and scenarios in a complete uncertainty assessment is not a trivial task. Design of experiment techniques such as the fractional factorial design may be used. This involves the following steps:

1. Prepare a complete list of geological and geostatistical factors.
2. Determine the possible range of variations of each factor.
3. Decrease the dimensionality of the problem by:
    a. Fast screening such as two-level fractional factorial design to determine the most important variables
    OR
    b. Aggregating the related variables together and decreasing the number of factors (see paper 112 of this report)
4. Perform three-level factorial design on the most important factors
5. Generate the response surface model (RSM)
6. Provide the best possible geostatistical realization for future reservoir simulation

The range of variation in the input factors can be obtained either from expert knowledge or statistical analysis of analogue data. One of the practical techniques for quantification of the uncertainty in the histogram is the spatial bootstrap method (Journel and Bitanov, 2004). This approach allows assessment of uncertainty by resampling original data accounting for their spatial correlation. The GSLIB type program called `Spatial_Bootstrap` (Deutsch, 2004) can be used to perform the spatial bootstrap between two correlated variables. Conditional Finite Domain (Babak and Deutsch, 2006) technique can also be used to capture the uncertainty in the input histogram. The spatial bootstrap can also be used to calculate the uncertainty of the experimental variogram at each lag separation distance (Derakhshan, 2007).

Incorporating parameter uncertainty in geostatistical simulation has not been well developed. Derakhshan and Deustch (2008) proposed a methodology to directly incorporate the uncertainty in mean in sequential Gaussian simulation based on the theory of generalized linear distributions. A GSLIB type program called `SGS-GLD` can be used to apply their methodology (Derakhshan, 2008). Alshehri (2010) presented a methodology to account for the parameter uncertainty in the mean for an application in quantification of uncertainty in structural surfaces.

4. Conclusion

Design of experiment has been used widely in engineering applications including reservoir engineering. A framework for the application of DOE in the quantification of uncertainty in geostatistical modeling is presented. Details of each step involved in the implementation of DOE in geostatistics was presented. A case study should be set up to implement the framework presented in this paper.

References

Alshehri, N.S., 2010, Quantification of reservoir uncertainty for optimal decision making, Ph.D. Thesis, University of Alberta, 185 p.

Amudo, C., Graf, T., Harris, N.R., Dandekar, F., May, R.S., 2008, Experimental Design and Response Surface Models as a Basis for Stochastic History Match – A Niger Delta Experience, International Petroleum Technology Conference, paper IPTC 12665, Malaysia.

Antony, J., 2003, Design of Experiments for Engineers and Scientists, Elsevier Science and Technology Books, ISBN: 0750647094, 152 p.

Babak, O. and Deutsch, C.V., 2006, A Conditional Finite-Domain (CFD) Approach to Parameter Uncertainty, *Centre for Computational Geostatistics*, paper 106, Report 8, University of Alberta.

Bu, T. and Damsleth, E., 1996, Errors and Uncertainties in Reservoir Performance Prediction, *SPEFE,* September, p. 194.

Chu, C.F., 1990, Prediction of Steamflood Performance in Heavy Oil Reservoirs Using Correlations Developed by Factorial Design Method, paper SPE 20020 presented at the 1990 California Regional Meeting, Ventura, California.

Damsleth, E., Hage, A., and Volden, R., 1992, Maximum Information at Minimum Cost: A North Sea Field Development Study with Experimental Design, *Journal of Petroleum Technology,* December, p. 1350.

Derakhshan, S.H., 2007, A decision Support System for Variogram Calculation, M.Sc. Thesis, University of Alberta, Canada, 75 p.

Derakhshan, S.H., 2008, Application of SGS-GLD to Parameter Uncertainty, *Centre for Computational Geostatistics*, paper 401, Report 10 , University of Alberta.

Derakhshan, S.H., and Deutsch, C.V., 2008, Generalized Linear Distributions to Account for Parameter Uncertainty in Sequential Gaussian Simulation, *Centre for Computational Geostatistics*, paper 104, Report 10 , University of Alberta.

Deutsch, C.V., 2002, Geostatistical Reservoir Modeling, *Oxford University Press*, New York, 376 p.

Deutsch, C.V., 2004, A Statistical Resampling Program for Correlated Data: Spatial_Bootstrap, *Centre for Computational Geostatistics*, paper 401, Report 6, University of Alberta.

Eide, A.L. *et al.,* 1994, Automatic History Matching by Use of Response Surfaces and Experimental Design, paper presented at the 1994 European Conference on the Mathematics of Oil Recovery, Roros, Norway, 7-10 June.

Journel, A.G. and Bitanov, A., 2004, *Uncertainty in N/G ratio in early reservoir development*, Journal of Petroleum Science and Engineering, 44 (1-2), pp.115-130.

Kjonsvik, D., Jacobsen, T., and Doyle, J., 1994, The Effects of Sedimentary Heterogeneities on Production From a Shallow Marine Reservoir-What Really Matters?, paper SPE 28455 presented at the 1994 European Petroleum Conference, London.

Leuangthong, O., 2005, Advances in Experimental Design Matrix for Establishing Sensitivity, *Centre for Computational Geostatistics*, paper 102, Report 7, University of Alberta.

Montgomery, D.C., 2001, Design and Analysis of Experiments, 5th edition, John Wiley and Sons Inc., New York, 684 p.

Narayanan, K. *et al.,* Response Surface Methods for Upscaling Heterogeneous Geologic Models, paper SPE 51923 presented at the 1999 SPE Reservoir Simulation Symposium, Houston.

Prada, J.W.V., and Cunha, L.B., 2008, Prediction of SAGD Performance Using Response Surface Correlations Developed by Experimental Design Techniques, *Journal of Canadian Petroleum Technology*, Vol. 47, No. 9, p. 58-64.

Prada, J.W.V., Deutsch, C.V., and Cunha, L.B., 2008, Uncertainty Assessment of SAGD Performance Using a Proxy Model Based on Butler's Theory, paper SPE 115662 presented at the 2008 SPE Annual Technical Conference and Exhibition, Denver, Colorado.

Willis, B.J. and White, C.D., 2000, Use of Quantitative Outcrop Description for Reservoir Modeling, *J. Sedimentary Research,* 70**,** No. 4, 788.