Toward a Reliable Quantification of Uncertainty on Production Forecasts: Adaptive Experimental Designs

C. Scheidt, I. Zabalza-Mezghani, M. Feraille, D. Collombier

▶ **To cite this version:**

C. Scheidt, I. Zabalza-Mezghani, M. Feraille, D. Collombier. Toward a Reliable Quantification of Uncertainty on Production Forecasts: Adaptive Experimental Designs. Oil & Gas Science and Technology - Revue d'IFP Energies nouvelles, 2007, 62 (2), pp.207-224. 10.2516/ogst:2007018 . hal-02005636

HAL Id: hal-02005636
https://ifp.hal.science/hal-02005636

Submitted on 4 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Oil & Gas Science and Technology – Rev. IFP, Vol. 62 (2007), No. 2, pp. 207-224
Copyright © 2007, Institut français du pétrole
DOI : 10.2516/ogst:2007018

IFP International Conference
Rencontres Scientifiques de l'IFP

Quantitative Methods in Reservoir Characterization
Méthodes quantitatives de caractérisation des réservoirs

Toward a Reliable Quantification of Uncertainty on Production Forecasts: Adaptive Experimental Designs

C. Scheidt[1], I. Zabalza-Mezghani[1], M. Feraille[1] and D. Collombier[2]

1 Institut Français du Pétrole, IFP, 1-4, avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex - France
2 Université Louis Pasteur, Strasbourg - France
e-mail: scheidtc@stanford.edu - isabelle.zabalza-mezghani@ifp.fr - mathieu.feraille@ifp.fr - collombier@math.u-strasbg.fr

Résumé — Vers une quantification fiable des incertitudes sur les estimations de production : plans d'expériences adaptatifs — La quantification des incertitudes est une phase essentielle dans l'évaluation des réservoirs pétroliers. La précision des estimations de production est fortement liée à l'incertitude sur les variables qui contrôlent les performances du réservoir (perméabilité, contact huile-eau, etc.). Le problème est complexe parce que l'effet des variables sur les performances du réservoir est souvent non-régulier, ce qui ne peut être détecté *a priori*.

La méthode des plans d'expériences est généralement utilisée pour quantifier les incertitudes sur la production et obtenir une représentation probabiliste de cette dernière, avec par exemple la détermination de scenarii de production P90, P50 et P10. En sélectionnant de manière optimale les simulations à effectuer, les plans d'expériences permettent la construction d'un modèle approché qui reproduit l'impact des paramètres incertains sur les performances du réservoir. L'utilisation de plan d'expériences permet de réaliser des analyses de risque tout en effectuant un nombre limité de simulations potentiellement très coûteuses en temps de calcul. Toutefois, l'utilisation des plans d'expériences est généralement associée à la construction de surfaces de réponses polynomiales de faible degré, elle montre ses limites dès lors que les paramètres incertains ont un impact non-régulier sur la réponse en production.

Nous présentons une nouvelle approche pour l'analyse de risque, fiable y compris lorsque l'impact des paramètres sur la réponse est non-régulier. Nous proposons de construire des plans d'expériences évolutifs, pour intégrer graduellement les non-régularités. Partant d'un plan d'expériences initial, la méthodologie détermine itérativement de nouvelles simulations susceptibles d'être informatives sur le comportement de la réponse. Inspirée de méthodes statistiques et de plans d'expériences, cette approche a montré son efficacité pour la modélisation de réponses complexes et non-régulières. Elle fournit une estimation fiable des incertitudes sur les estimations de production.

Abstract — Toward a Reliable Quantification of Uncertainty on Production Forecasts: Adaptive Experimental Designs — *Quantification of uncertainty in reservoir performance is an essential phase of proper field evaluation. The reliability of reservoir forecasts is strongly linked to the uncertainty in the information we have about the variables that control reservoir performance (*e.g. *permeability, oil-water contact, etc.). The problem is complex, since the effect of the variables on the reservoir performance is often non-linear, which cannot be inferred* a priori.

Experimental design methods are well-known and widely used to quantify uncertainty and obtain probabilistic representation of production through, for instance, the P90, P50 and P10 production scenarios. By optimally selecting the flow simulations that should be performed, experimental design builds a proxy model that mimics the impact of the uncertain parameters on the reservoir performance. Using experimental design, one can perform risk assessment while performing a limited number of potentially expensive fluid flow simulation runs. However, experimental designs are based on simple polynomial response surface approximations, which show clearly their limits when the production response varies irregularly with respect to reservoir parameters.

We present a new approach to properly assess risk even if the impact of the uncertain parameters is highly irregular. Contrary to classical experimental designs which assume a regular, 1st or 2nd degree polynomial-type behavior of the response, we propose to build evolutive designs, to fit gradually the potentially irregular shape of the uncertainty. Starting from an initial trend of the uncertainty behavior, the method determines iteratively new simulations that might bring crucial new information to update the current estimation of the uncertainty. Inspired by statistical methods and experimental designs, this original methodology has demonstrated its efficiency in modeling accurately complex, irregular responses, and thus in providing reliable uncertainty estimation on production forecasts.

INTRODUCTION

Modeling of oil and gas fields has become increasingly sophisticated during the last decades. With the advent of high technology data acquisition programs, it is now possible to calibrate complex numerical models that integrate a large degree of information concerning reservoir behavior. However, the data acquisition, its processing, interpretation and integration, can have significant errors. These errors, combined with the necessary geological assumptions, lead to a complex reservoir model which is populated by a very large number of uncertain parameters. The main consequence is that the output of the reservoir modeling workflow, for instance, fluid in place, reserves and production profiles are uncertain.

In order to have realistic production forecasts, it is essential to take into account the characteristics of the uncertain parameters influencing the production response. This is usually done in two steps. In the first step, it is essential to identify, within all the specified uncertain parameters, those parameters which have a large influence on production forecasts. This is termed a sensibility study. Then in the second step, a quantification of the risk associated to the influential parameters is performed. From the results of this step, the reservoir engineer is able to make important decisions during reservoir exploitation, while taking the parameter uncertainty into account.

Quantifying the impact of each uncertain parameter on production requires many fluid flow simulations. Because of the large number of parameters and the physical complexity of the reservoir, fluid flow models are difficult to construct and very computationally time consuming. In order to control the cost of an uncertainty study, classical experimental design methods are commonly used to create proxy models of the production response. These proxy models can be a good representation of the fluid flow simulator. Their evaluation is very rapid, and therefore allows for risk analysis using Monte Carlo sampling. Constructing a proxy model is therefore essential to obtain reliable production forecasts which account for uncertainty. A proxy model needs to satisfy the two following requirements:

– to reproduce the behavior of the simulated data: the model should fit the simulations already performed using the fluid flow simulator;
– to predict reliably the behavior of the production response for any value of the parameters where no simulations had been performed.

Recent work [1, 2] has been performed on the subject of uncertainty management in reservoir engineering. These authors construct 1st or 2nd order polynomial proxy models using classical experimental designs. However, this approach is limited to regular responses, *i.e.* for responses which can be approximated with a low order polynomial model. However, this is not correct for many reservoir engineering applications, as illustrated in Figure 1. Therefore, in order to estimate uncertainty, it is essential to develop a methodology, which can reliably predict irregular responses, but requires few simulation runs. A good compromise between the computational cost of the study and the model quality must be reached.

In this paper, we propose a methodology to construct an effective proxy model of the fluid flow simulator which takes into account the potential irregularities of the response. Contrary to classical experimental designs which assume a polynomial behavior of the response, we propose a procedure to build evolutive experimental designs to fit gradually the

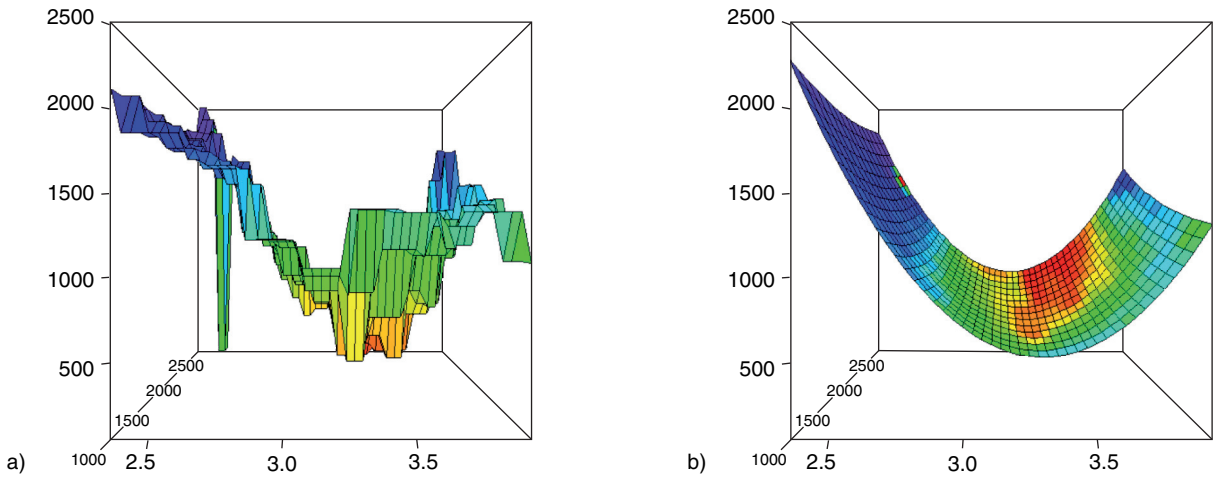

Figure 1

(a) Representation of an irregular production response.
(b) 2nd order polynomial model for this response [13].

potentially irregular shape of uncertainty (as shown in *Fig. 1a*). This methodology benefits from the advantage of experimental design, which allows the control of the number of fluid flow simulations, combined with the flexibility to study irregular behavior which is too complex for classical statistical methodologies. The evolutive approach also does not require assumptions as to the shape of the response, which is difficult to quantify *a priori*.

The outline of the paper is as follow. In Section 1, the principle of the evolutive methodology is explained. We then present the proxy model used to fit the data, which is a combination of regression and residual kriging. In Sections 2 and 3, we propose criteria to enhance iteratively the initial experimental design, starting with the search of stationary points of the response and then the search in under-sampled area of the experimental design. Sections 4 and 5 present validation tests on analytical functions and a real field case from offshore Brazil. We then conclude with a discussion on possible directions of future work.

1 ADAPTIVE MODELING AND EVOLUTIVE EXPERIMENTAL DESIGNS

1.1 Principle of the Methodology

This methodology aims at modeling an irregular function f, for example the results of a fluid flow simulator, depending on k uncertain reservoir parameters, with a reasonable number of simulations. The simulations used for building the initial proxy model are located employing experimental design techniques. Let us consider the following notations:

- k, the number of uncertain parameters and n the current number of simulations performed;
- $f: \mathbf{R}^k \rightarrow \mathbf{R}$ the fluid flow simulator model;
- $X_j, j=1,..., k$, the uncertain parameters;
- $\boldsymbol{x}_i$, the locations of the performed simulations: $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{ik}), i = 1, ..., n$;
- $f(\boldsymbol{x}_i)$, the simulated value obtained at the location $\boldsymbol{x}_i$: $\boldsymbol{f} = (f(\boldsymbol{x}_i))_{i=1,...,n}$;

$D = \prod_{j=1}^{k} [X_j^{\min}, X_j^{\max}]$, the experimental domain with $X_j^{\min}$ and $X_j^{\max}$ respectively the minimal and maximal value that can have the j-th parameter.

The methodology is based on an association of two methods: evolutive experimental designs and kriging estimation. We propose an adaptive technique in order to capture, step by step, the irregularity of the response. An illustration of the principle of the methodology is presented in Figure 2.

The first step of the methodology is to construct an initial experimental design, noted $X_{(1)}$. Since reservoir simulation can be time consuming, it is important to select a small number of simulations which provides a maximum of information. Experimental design is used for selecting an optimal set of simulations to obtain an initial approximation (proxy model) of the response behavior as a function of the influential uncertain parameters. In the current context, in the absence of prior knowledge about the response behavior, we propose the use of Latin hypercubes with maximin distance. Maximin Latin hypercubes rely on a uniform spatial distribution of the points which maximize the minimum distance between two points [3]. This kind of design provides a good starting point to capture the mean behavior of the response. Reservoir simulations are performed following this initial

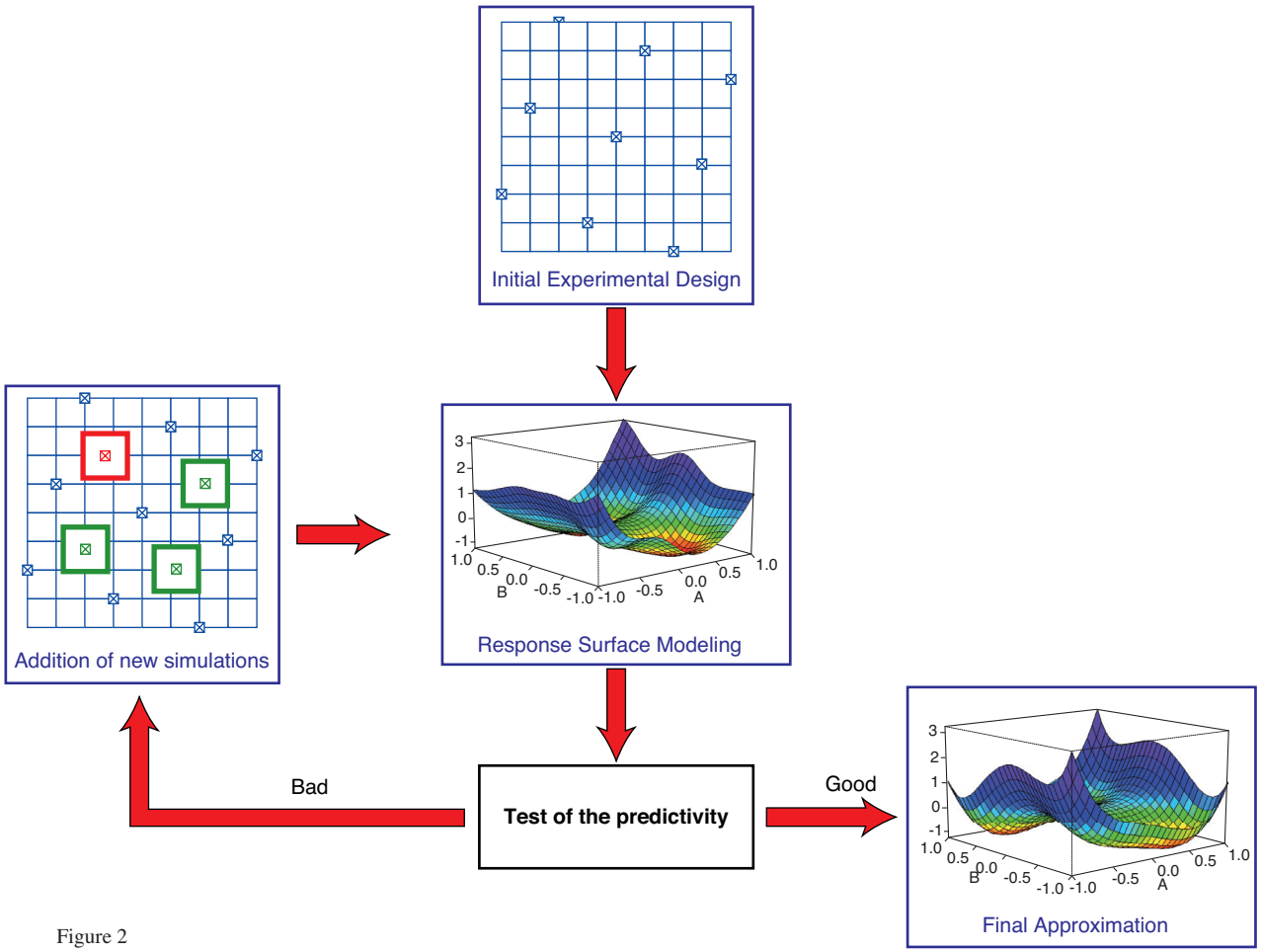

Figure 2

Principle of the adaptive methodology.

experimental design $X_{(1)}$ to obtain observations of the production response and to fit a first approximation $\tilde{f}_{(1)}$. At this point, the methodology aims at determining iteratively new informative simulations that are added to the experimental design $X_{(i)}$. The location of new simulation runs is optimized according to the current experimental design $X_{(i)}$ and approximation $\tilde{f}_{(i)}$ in order to maximize the amount of information brought by each simulation.

We thus obtain a set of hierarchical experimental designs $X_{(i)}$ $i = 1,..., m$ such as $X_{(1)} \subset X_{(2)} \subset X_{(3)} \subset ... \subset X_{(m)} \in D$, and a set of corresponding approximations $\tilde{f}_{(i)}$ $i = 1,..., m$. The first approximation $\tilde{f}_{(1)}$ is obtained on the basis of the initial experimental design $X_{(1)}$, which includes few data, therefore may not be very accurate. As the number of iterations i increases, the experimental design $X_{(i)}$ contains more simulations and the corresponding approximation $\tilde{f}_{(i)}$ becomes gradually more reliable.

One strength of this approach is that it allows the user to control the number of simulations based upon desired accuracy of the current approximation: at each iteration i, one can decide either to add simulations to get a more accurate approximation $\tilde{f}_{(i+1)}$, or to sacrifice approximation accuracy in order to reduce the overall simulation time and consider $\tilde{f}_{(i)}$ as the final proxy model of the fluid flow simulator.

1.2 Definition of the Proxy Model

In the case of computer experiments, the data are deterministic, *i.e.* running the simulator with the same parameters values gives identical observations. Therefore, the response surface should exactly fit the simulator data. For this reason, interpolation techniques are usually preferred to regression models for computer experiments. The basic idea, suggested initially by Sacks *et al.* [4, 5], consists in adding an error term to the polynomial approximation of the trend of the observations. To do so, we assume that the proxy model is defined as the sum of a regression model and a random process:

$$\tilde{f}(\boldsymbol{x}) = \sum_{j}^{p} g_j(\boldsymbol{x})\beta_j + Z(\boldsymbol{x}) = G(\boldsymbol{x})\beta + Z(\boldsymbol{x})$$

with:

– $G(\boldsymbol{x})\ \beta = \mathrm{E}\left[f(\boldsymbol{x})\right]$ the deterministic part of the model, for instance a 2nd order polynomial. The coefficients $g_1(.),..., g_p(.)$ are known regression functions and $\boldsymbol{\beta}=(\beta_1,..., \beta_p)$ is the vector of unknown regression coefficients.

– $Z(\boldsymbol{x})$, the residuals of the response $\boldsymbol{f}$ with respect to the deterministic approximation. Z is assumed to be a stationary random process with zero mean and covariance between $Z(\boldsymbol{t})$ and $Z(\boldsymbol{u})$ defined by: $C(\boldsymbol{t},\boldsymbol{u}) = Cov\left(Z(\boldsymbol{t}), Z(\boldsymbol{u})\right) = \sigma^2 R(\boldsymbol{t},\boldsymbol{u})$, where σ^2 is the variance of the process and $R(\boldsymbol{t},\boldsymbol{u})$ the correlation function between $Z(\boldsymbol{t})$ and $Z(\boldsymbol{u})$.

The choice of a low order polynomial model to adjust the trend of the response is driven by the desire to limit the number of required simulations to be performed for this initial step. This initial fitting of the response is obtained using a traditional least-square optimization. Note that the approach is valid for any kind of initial approximation model, however complex its structure.

The residual model is obtained by kriging. We assume that the least squares optimization leads to an unbiased estimate of the trend of the response. In this case, the residuals have no mean, and simple kriging is sufficient for modeling the residual. However, if the trend estimator is biased, ordinary kriging should be used. The kriging predictor at $\boldsymbol{x}_0$, in the case of simple kriging with zero mean, is [6]:

$$\hat{Z}\left(\boldsymbol{x}_0\right) = {}^t\boldsymbol{c}\left(\boldsymbol{x}_0\right)\boldsymbol{C}^{-1}\boldsymbol{Z}$$

with:

– ${}^t\boldsymbol{c}(\boldsymbol{x}_0) = \left[Cov\left(Z(\boldsymbol{x}_0), Z(\boldsymbol{x}_l)\right), l = 1, ..., n\right]$, the covariance of Z between $\boldsymbol{x}_0$ and the points of the current experimental design $X_{(i)}$,

– $\boldsymbol{C} = \left[Cov\left(Z(\boldsymbol{x}_l), Z(\boldsymbol{x}_j)\right), l, j = 1, ..., n\right]$, the covariance of Z, for instance exponential or Gaussian.

In our adaptive approach, we use universal kriging to model the production response, its expression is:

$$\tilde{f}(\boldsymbol{x}_0) = \boldsymbol{G}(\boldsymbol{x}_0)\,\hat{\boldsymbol{\beta}} + {}^t c(\boldsymbol{x}_0)\boldsymbol{C}^{-1}\left[\boldsymbol{f} - \boldsymbol{G}\hat{\boldsymbol{\beta}}\right] \qquad (1)$$

Note that Equation (1) is a function of the covariance between the observations. In the case of computer experiments, we usually have too few simulations to fit a covariance (or variogram) model. In general, the spatial correlations of a response are extremely difficult to infer, and no prior information about the structure of the covariance is available. We employ two methods to fit a variogram depending on the quantity of simulated data. A detailed description of the methods is provided in Appendix A.

The initial experimental design $X_{(1)}$ includes few simulation runs. If the predictivity of the model is poor, new simulations must be added to the experimental design to explore new areas and to improve the accuracy of the approximation. In this paper, we propose three criteria of adding new points, divided into two categories: search of the stationary points of the current approximation $\tilde{f}_{(i)}$, and a more global search, consisting in sampling areas where no simulations have been performed.

2 RESPONSE SURFACE MODELING: CAPTURING OF STATIONARY POINTS

2.1 Search for Maximum and Minimum Values

The first criterion for adding new simulation runs attempts to find the extremes of the response. Therefore, we look for the maxima and minima of the current approximation $\tilde{f}_{(i)}$, in order to improve the current experimental design $X_{(i)}$. We present here the methodology for the search of the maxima. The same principle is used for the minima.

The production response is assumed to be irregular, so we look for several maxima using local optimization with different starting points $\boldsymbol{x}_{init}^j$. The search is realized in a neighborhood ϑ_j of the starting points as illustrated in Figure 3. In our case, the objective is not to find the global maximum of the response. The local maxima are also of interest because they allow for a better characterization of the overall response surface.

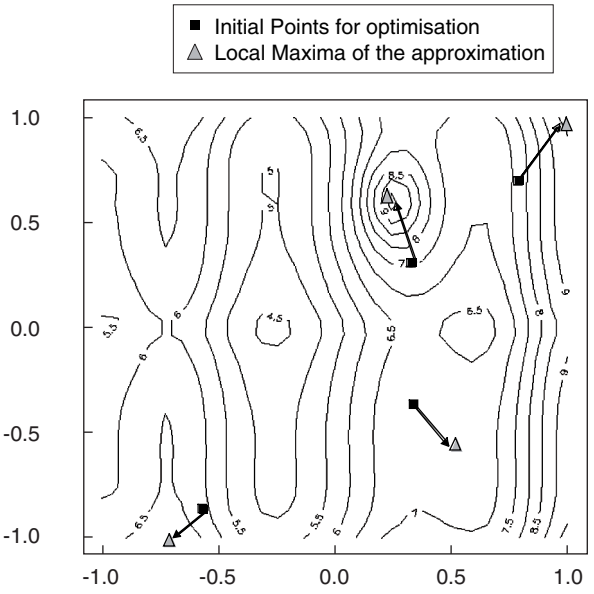

Figure 3

Maximization for several starting initial points.

One local maximum of the current approximation (proxy model) $\tilde{f}_{(i)}$ in a neighborhood ϑ_j of $\boldsymbol{x}_{init}^j$ is given by:

$$\boldsymbol{x}_{max}^{\vartheta_j} = \arg\max_{\boldsymbol{x}\in\vartheta_j}\left(\tilde{f}_{(i)}(\boldsymbol{x})\right) \text{ where } \vartheta_j \text{ is a neighborhood of } \boldsymbol{x}_{init}^j$$

A flow simulation is subsequently performed at $\boldsymbol{x}_{max}^{\vartheta_j}$ and a new proxy $\tilde{f}_{(i+1)}$ is fitted. Because this initial search of maximum is based on the approximation $\tilde{f}_{(i)}$, a test is

performed: we compare the value of the response at the point $\boldsymbol{x}_{max}^{\vartheta_j}$ with the value of the response at the point $\boldsymbol{x}_{init}^{j}$:

$$\Delta_{rep}^{\vartheta_j} = \tilde{f}_{(i+1)}(\boldsymbol{x}_{max}^{\vartheta_j}) - \tilde{f}_{(i+1)}(\boldsymbol{x}_{init}^{j})$$

- if $\Delta_{rep}^{\vartheta_j}$ is small, we assume that $\boldsymbol{x}_{max}^{\vartheta_j}$ is close to a maximum of the current approximation $\tilde{f}_{(i+1)}$ (the response has not increased sufficiently) - no more information is needed in the area of $\boldsymbol{x}_{init}^{j}$,
- if $\Delta_{rep}^{\vartheta_j}$ is negative, the assumption of a local maximum in the neighborhood ϑ_j is invalid - no more simulations are performed,
- if $\Delta_{rep}^{\vartheta_j}$ is large, the addition of $\boldsymbol{x}_{max}^{\vartheta_j}$ in the neighborhood of $\boldsymbol{x}_{init}^{j}$ is valid - perform a line search to simulate a new point in the direction of increasing $\tilde{f}_{(i+1)}$ in order to find the maximum.

In the later case, since we have two points in ϑ_j, the line search for a new maximum is inspired by the simplex method proposed by Nelder and Mead [7]. The coordinates of the new point $\boldsymbol{x}_{new}^{\vartheta_j}$ in the increasing direction are:

$$\boldsymbol{x}_{new}^{\vartheta_j} = \boldsymbol{x}_{init}^{j} + (\boldsymbol{x}_{init}^{j} - \boldsymbol{x}_{max}^{\vartheta_j})$$

A simulation is then performed at the point $\boldsymbol{x}_{new}^{\vartheta_j}$, and once again, we calculate the amount of information brought by the simulation, *i.e.* we calculate $\Delta_{rep}^{\vartheta_j}$. This iterative procedure continues until the simulation of the new point is not sufficiently informative.

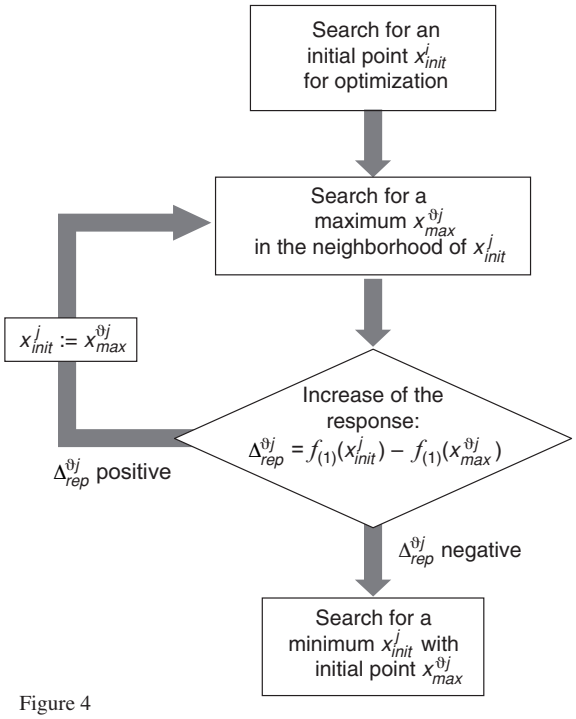

Figure 4

Iterative search of maximum.

The aim of this method is to find the extremes of the response. In order to best characterize the local extremes and the sinuosity of the response, we propose to search for minima starting from maxima determined previously. Therefore, we search for the point $\boldsymbol{x}_{init}^{j}$, with initial point $\boldsymbol{x}_{new}^{\vartheta_j}$:

$\boldsymbol{x}_{min}^{j} = \arg\min_{\boldsymbol{x}\in D} \tilde{f}_{(i)}(\boldsymbol{x})$, where D is the experimental domain.

The search for extremes is illustrated in Figure 4. Note that this method makes use of the knowledge of the current approximation and the values of the response in the explored neighborhood.

This approach is crucial in the adaptive method, since it allows the characterization of the sinuosity and the variability of the response. However, this criterion alone does not catch all the irregularities of the response. Therefore, we propose to search for points where one or more partial derivatives are zero.

2.2 Search for Null Derivative Points

In order to better characterize the variations of the response, we propose to search for points where one or more partial derivatives are zero. The dependence of the response with respect to the parameters is not isotropic. Thus, we propose to perform at each iteration a flow simulation at points for which at least one of the k partial derivative is zero, where k is the number of parameters.

The derivative of the estimator $\tilde{f}_{(i)}$ of the response at each point $\boldsymbol{x}$ of the experimental domain is given by:

$$\frac{\partial \tilde{f}_{(i)}}{\partial X_j}(\boldsymbol{x}) = \frac{\partial \boldsymbol{G}}{\partial X_j}(\boldsymbol{x})\hat{\beta} + \frac{\partial \boldsymbol{c}^t}{\partial X_j}(\boldsymbol{x})\boldsymbol{C}^{-1}\left[\boldsymbol{f} - \boldsymbol{G}\hat{\beta}\right], \quad j = 1,..,k$$

The search of the zero of $\dfrac{\partial \tilde{f}_{(i)}}{\partial X_j}$ can be written as a function to minimize:

$$\boldsymbol{x}_g = \left\{ \boldsymbol{x}, \exists\, X_j, i = 1,\ldots,k, \quad \text{where} \quad \boldsymbol{x} = \arg\min_{\boldsymbol{x}\in D} \left(\frac{\partial \tilde{f}_{(i)}(\boldsymbol{x})}{\partial \boldsymbol{X}_j} \right)^2 \right\}$$

As in the previous procedure, the minimization is performed using local optimization at several initial points. As the current approximation increases in accuracy, the simulation at point $\boldsymbol{x}_g$ should allow the determination of extremes or directional inflexion points of the response. At each iteration the point $\boldsymbol{x}_g$ is determined based on the current approximation which may be inaccurate. The point $\boldsymbol{x}_g$ is considered to be a stationary point in the direction defined by X_j. To confirm this assumption, we construct a new approximation $\tilde{f}_{(i+1)}$ using Equation (1) which incorporates the point $\boldsymbol{x}_g$, and we calculate the numerical directional derivative at $\boldsymbol{x}_g$:

$$\tilde{f}'^{j}_{(i+1)}(\boldsymbol{x}_g) = \lim_{\varepsilon \to 0} \frac{\tilde{f}_{(i+1)}(\boldsymbol{x}_g + \varepsilon) - \tilde{f}_{(i+1)}(\boldsymbol{x}_g)}{\varepsilon}$$

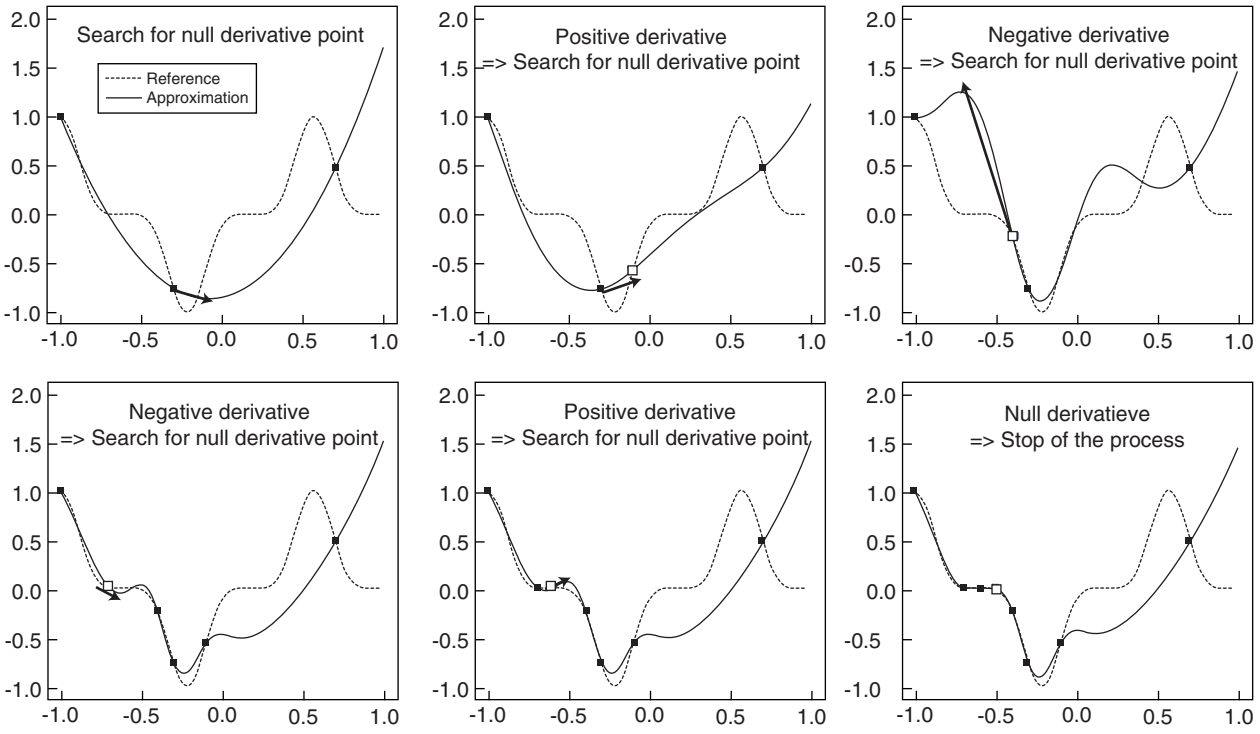

Figure 5

Iterative search of an inflexion point.

where $\varepsilon = (0,...,0,\varepsilon_j,0,...,0)$.

A test of the value of $\tilde{f}'_{(i+1)}(\boldsymbol{x}_g)$ is then performed:

– if $\tilde{f}'_{(i+1)}(\boldsymbol{x}_g) \approx 0$, the search of stationary points at the current iteration i is accomplished with the determination of $\boldsymbol{x}_g$,
– if $\tilde{f}'_{(i+1)}(\boldsymbol{x}_g) \neq 0$, the point is not stationary for the new approximation. A new search for a stationary point is performed, starting from $\boldsymbol{x}_g$.

Note that it is necessary here to use the numerical derivative because the analytical derivative is not defined at the points of the experimental design.

An example of how null partial derivative points are added is presented in Figure 5. In this example, one null partial derivative point is added as long as the value of the numerical derivative does not equal zero. We can see that an inflexion point is detected.

The search for points of null partial derivative allows for the detection of the inflexion points, thus capturing the curvature of the response.

Both criteria presented, the search of the extremes and the directional null derivative points, are important for the modeling of irregular responses. Later, in Section 4.2, we will demonstrate the importance of each search criterion. Note however that since they are based on the current approximation of the response, they can lead to over-sampling of regions already explored without sampling in new areas for which little or no information is available. In order to sample areas of sparse information, we study, in Section 3, the behavior of the current approximation at points where the kriging variance is maximal. These points are the most distant from points already simulated in the experimental domain.

3 SURFACE GLOBAL FITTING: UNDER-SAMPLED AREAS AND PILOT POINTS

3.1 Maximum Kriging Variance

One advantage of the use of universal kriging for the proxy model is that it provides information on the kriging variance. As illustrated in Figure 6, the construction of such an estimator implies that the kriging variance is equal to zero at the experimental design points and maximal in the under-sampled areas.

A large kriging variance indicates a lack of information that would constrain more accurately the response surface model. Therefore, the points which maximize the kriging variance are good candidates for performing flow simulation: they are located in areas where no information is provided by the current experimental design. Consequently,

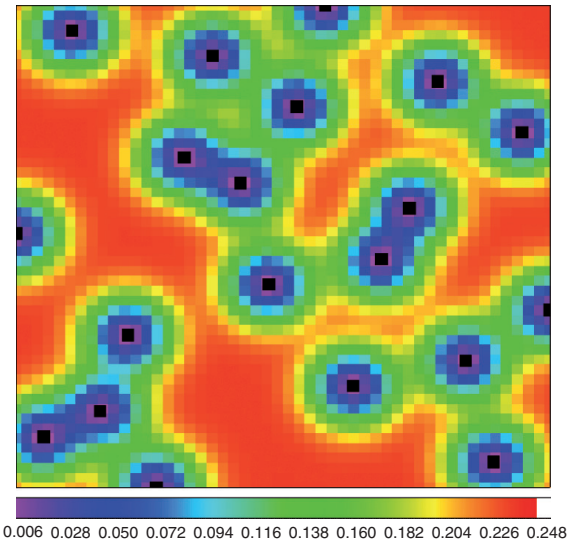

Figure 6

Kriging variance.

we search for the points $\boldsymbol{x}$ which maximize the kriging variance, *i.e.* we consider the criteria L:

$$L = \max_{\boldsymbol{x}\in D} \left\{ \tilde{\sigma}^2 \left(1 - {}^t\boldsymbol{r}(\boldsymbol{x})\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x})\right)\right\}$$

where:

– $\tilde{\sigma}^2$ is usually estimated by maximum likelihood method, as proposed initially by Currin *et al.* [8] and recommended later by many authors [9, 10]:

$$\tilde{\sigma}^2 = \tilde{\sigma}^2_{MLE}(\theta) = \frac{1}{n-q}\,{}^t(\boldsymbol{f} - \boldsymbol{G}\hat{\beta})\boldsymbol{R}^{-1}(\boldsymbol{f} - \boldsymbol{G}\hat{\beta})$$

– $r(\boldsymbol{x})$ the correlation between the points of the current experimental $X_{(i)}$ and the point $\boldsymbol{x}$,

– $\boldsymbol{R}$ the correlation between the points of the current experimental $X_{(i)}$,

– n the size of the current experimental $X_{(i)}$ and q = range ($\boldsymbol{G}$).

A natural approach could consist in directly simulating the point of maximal kriging variance. This procedure ensures the sampling of unexplored areas of the experimental domain. However, continual repetition of this procedure would eventually be equivalent to simulation using a uniform design, which would be too computationally time consuming, and would result in a poor sampling of the parameter space. Instead of using this uniquely spatial criterion of maximizing the kriging variance, in our approach, we propose to integrate information concerning the behavior of the current approximation of the response. To do this, we test the sensitivity of the approximation to new information at the points of maximum kriging variance, and then perform simulations only at the most sensitive points. This test is performed using the concept of pilot points.

3.2 Pilot Points

The pilot point methodology was initially introduced by De Marsily in 1995 [11]. It consists in using points for which no simulations are performed, whose values are assigned and varied according to the pilot point algorithm. In our application, we use pilot points to help to determine the most sensitive areas of the uncertain domain for which a flow simulation is necessary. Our strategy is divided in two main steps:

– determination of the pilot points $\boldsymbol{x}_i^{PP}$ by searching for points of maximum kriging variance;

– study of the sensitivity of the current approximation to new information at the points $\boldsymbol{x}_i^{PP}$.

In order to test the impact of the point $\boldsymbol{x}_i^{PP}$ on the approximation, we attribute a fictitious value Y_i^{PP} at $\boldsymbol{x}_i^{PP}$, then modify the value of Y_i^{PP} within an interval I_i^{PP}, and finally evaluate the effect on the classical predictivity coefficient Q_2 [12]. Through this procedure, we can identify the points for which additional information (obtained by flow simulation) may considerably change the shape of the approximation, particularly in terms of predictivity.

Due to the structure of the kriging model $\tilde{f}$, the classical prediction interval is given by [13]:

$$Y_i^{PP} \in I_i^{PP} = \left[\begin{array}{l} \tilde{f}(\boldsymbol{x}_i^{PP}) - q_\alpha \sqrt{\tilde{\sigma}^2 \left(1 - {}^t\boldsymbol{r}(\boldsymbol{x}_i^{PP})\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}_i^{PP})\right)}; \\ \tilde{f}(\boldsymbol{x}_i^{PP}) + q_\alpha \sqrt{\tilde{\sigma}^2 \left(1 - {}^t\boldsymbol{r}(\boldsymbol{x}_i^{PP})\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}_i^{PP})\right)} \end{array} \right]$$

with the q_α quantiles of the pivotal distribution of Z.

In order to select the most sensitive points to new information and ensure a sufficient increase in the predictivity of the approximation, we evaluate the value of Q_2 by varying the values of $Y_i^{PP} \in I_i^{PP}$. Depending on the variation of Q_2, we perform:

– Flow simulation at the points x_i^{PP} for which the variation of Q_2 is maximal and Q_2 is large. In this case, the addition of real information is warranted in order to stabilize the predictivity;

– Suppression of the pilot points which induce no variation of the predictivity, or result in a low value of predictivity. If Q_2 is small, we prefer to continue the iterative process by the addition of stationary points. If the variation of the value of the pilot point does not imply a variation of Q_2, then the simulation of this point is not considered informative.

The principle of the pilot point method used to test the predictivity of the approximation on different points is illustrated in Figure 7.

To provide an example, Figure 8 illustrates the case of 3 pilot points, located at the maximum of the kriging variance. In Figure 8, the top graph presents the location of the pilot points and their associated interval of fictitious values,

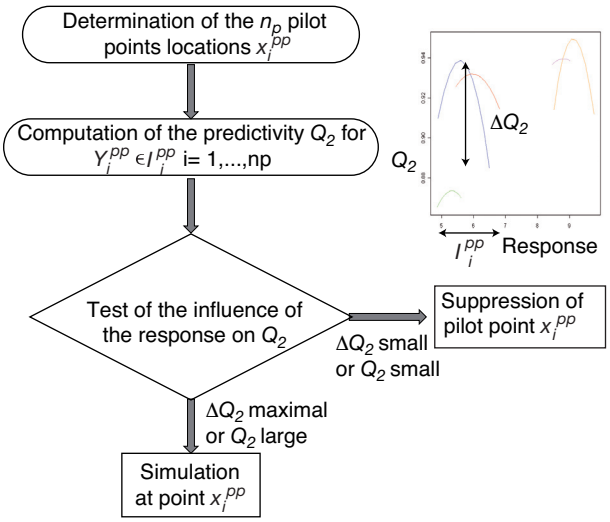

Figure 7

Workflow of the pilot point methodology to test the sensitivity of the approximation on predictivity.

shown with red arrows. The bottom three graphs show the variation of the Q_2 coefficient at each $\boldsymbol{x}_i^{PP}$ with respect to the pilot point value Y_i^{PP}. In this case, according to the variation of the Q_2 coefficient, we eliminate the first point of the experimental design (shown in the bottom left graph), and we perform flow simulation at the second and third points (bottom middle and bottom right graphs).

Note that the pilot point technique can also be used to optimize the predictivity of the current approximation $\tilde{f}$. In the optimization procedure, the fictitious values at the pilot points are modified in order to maximize Q_2. This idea is discussed in greater detail in Section 6.

3.3 Summary

Three criteria have been proposed above for adding new simulation points to the current experimental design. The goal is to optimize the location of the simulation points which then bring a maximum of information to the current approximation. The optimization procedure is illustrated in Figure 9.

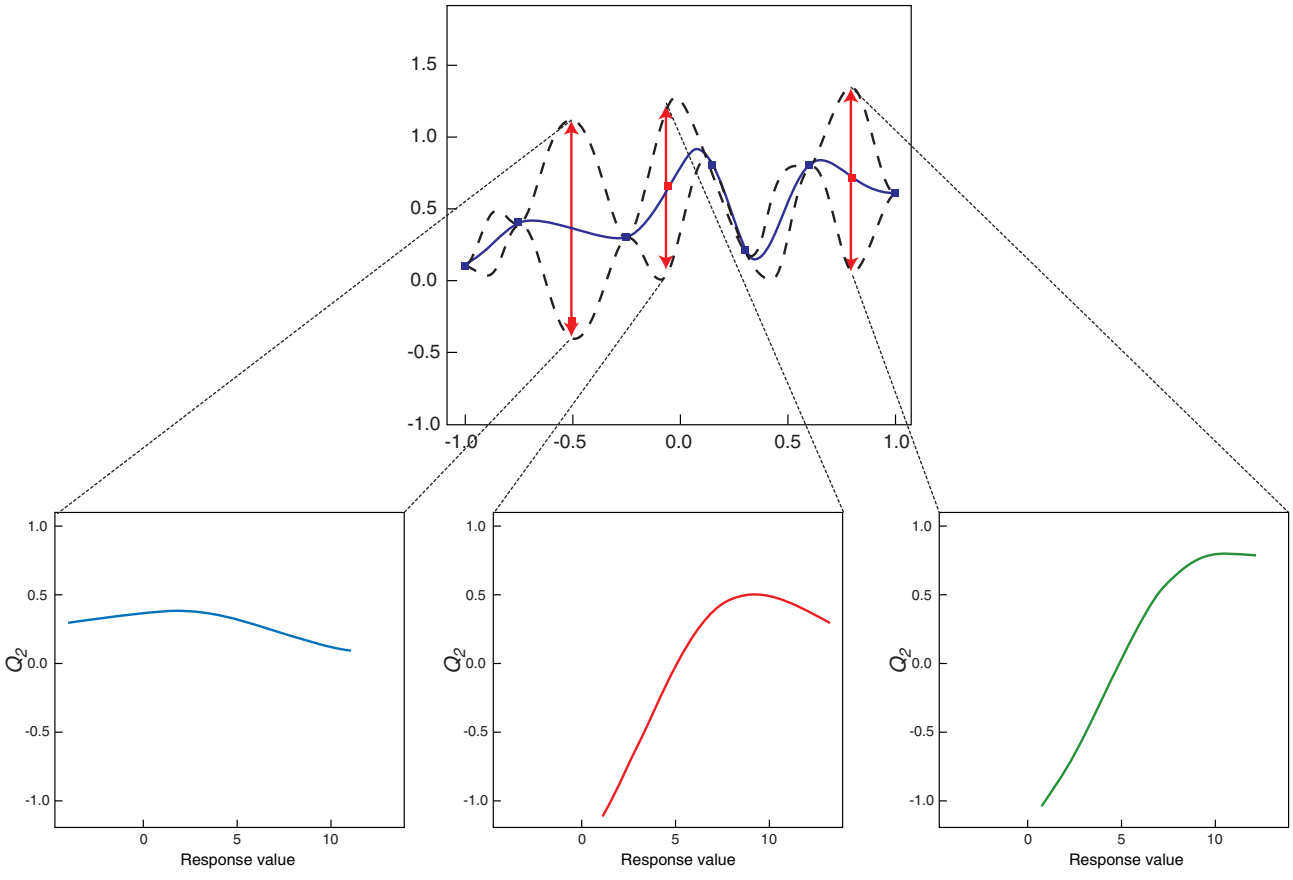

Figure 8

Estimation of the predictivity coefficient Q_2 as a function of the fictitious values attributed at each pilot points.

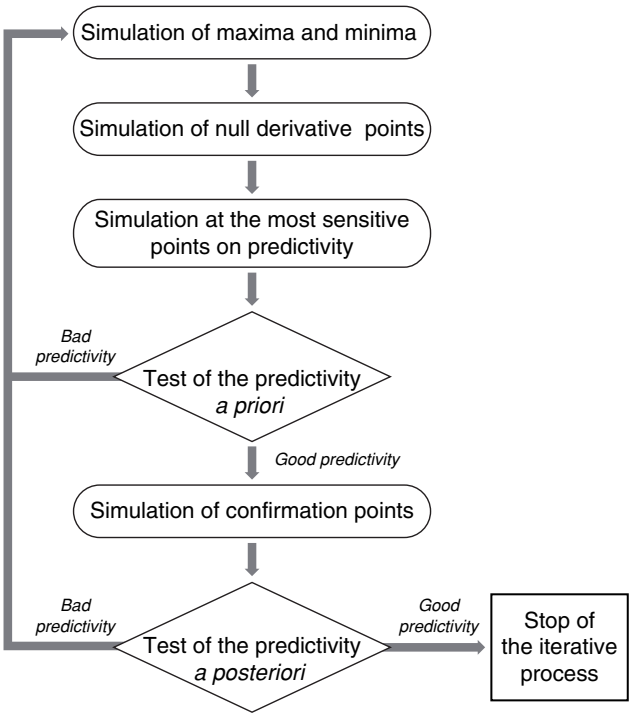

Figure 9

Workflow of addition of new simulations.

The confirmation points are selected at the points of maximum kriging variance, *i.e.* where sparse or no information is available. The purpose of the confirmation simulations is to test the posterior predictivity of the approximation, *i.e.* based upon the accuracy of the current model to the new points in the uncertain domain, the user can decide if more iterations are necessary, or if current approximation is sufficiently accurate.

4 VALIDATION OF METHODOLOGY USING ANALYTICAL FUNCTIONS

4.1 Modeling of a 2D Irregular Function

We present here an example of validation of our approach, which is tested using an explicit irregular analytical function, given by:

$$F(x,y) = 0.2e^{(x-3)} + 2.2|y| + 1.3y^6 - 2y^2 - 0.5y^4 - 0.5x^4 + 2.5x^2 \cdot$$
$$+ \ 0.7x^3 \ + \ \frac{3}{(8x-2)^2 + (5y-3)^2 + 1} + \sin(5x)\cos(3x^2)$$

Using the adaptive methodology, we propose in this example to simulate, at each iteration, 1 maximum and 1 minimum, 3 null derivative points and 3 points of maximum kriging variance through pilot points. The covariance is assumed to be Gaussian and anisotropic.

In Figure 10, we present the results of the modeling of function F within the domain $[-1,1]^2$. In Figure 10a, we show the 3D view of the function. Figure 10b presents the iso-line plots of the reference (in black), and the approximation (in blue) showing the simulation points in the experimental design. Finally, Figure 10c plots the predicted values obtained with the approximation versus the simulated values on a grid of 900 points. In this latter plot, we present the relative error, which is given by:

$$\varepsilon_{rel} = \frac{\sum_{i=1}^{n} \left| f(\boldsymbol{x}_i) - \tilde{f}(\boldsymbol{x}_i) \right|}{\sum_{i=1}^{n} \left| f(\boldsymbol{x}_i) \right|}$$

After a total number of 58 simulations, the approximation of the function F using the adaptive modeling is of good

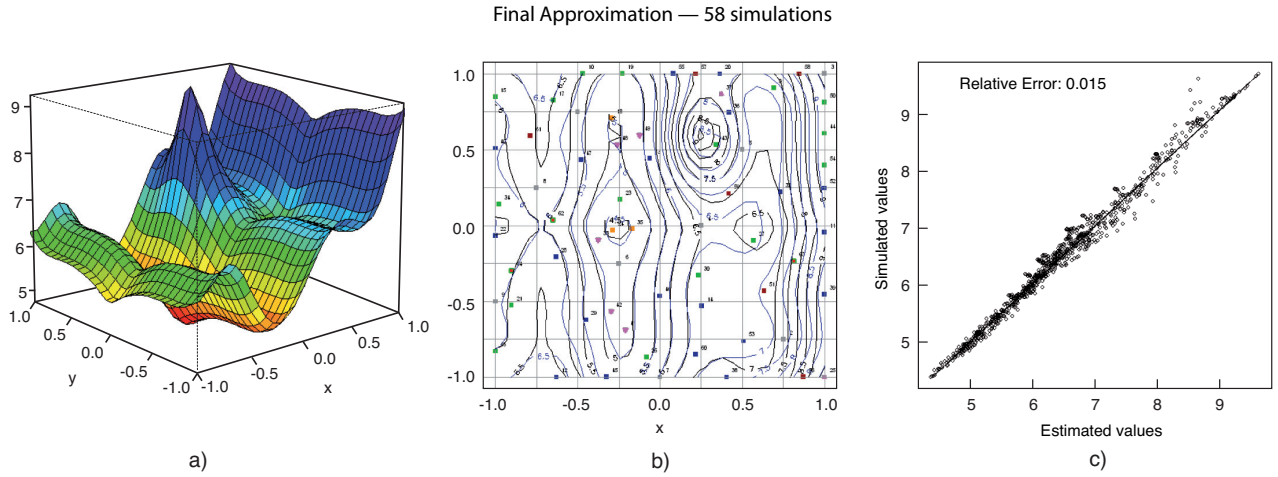

Figure 10

Modeling of irregular response: (a) perspective view of the reference; (b) iso-lines of the reference and approximation + experimental design; (c) simulated value versus predicted values.

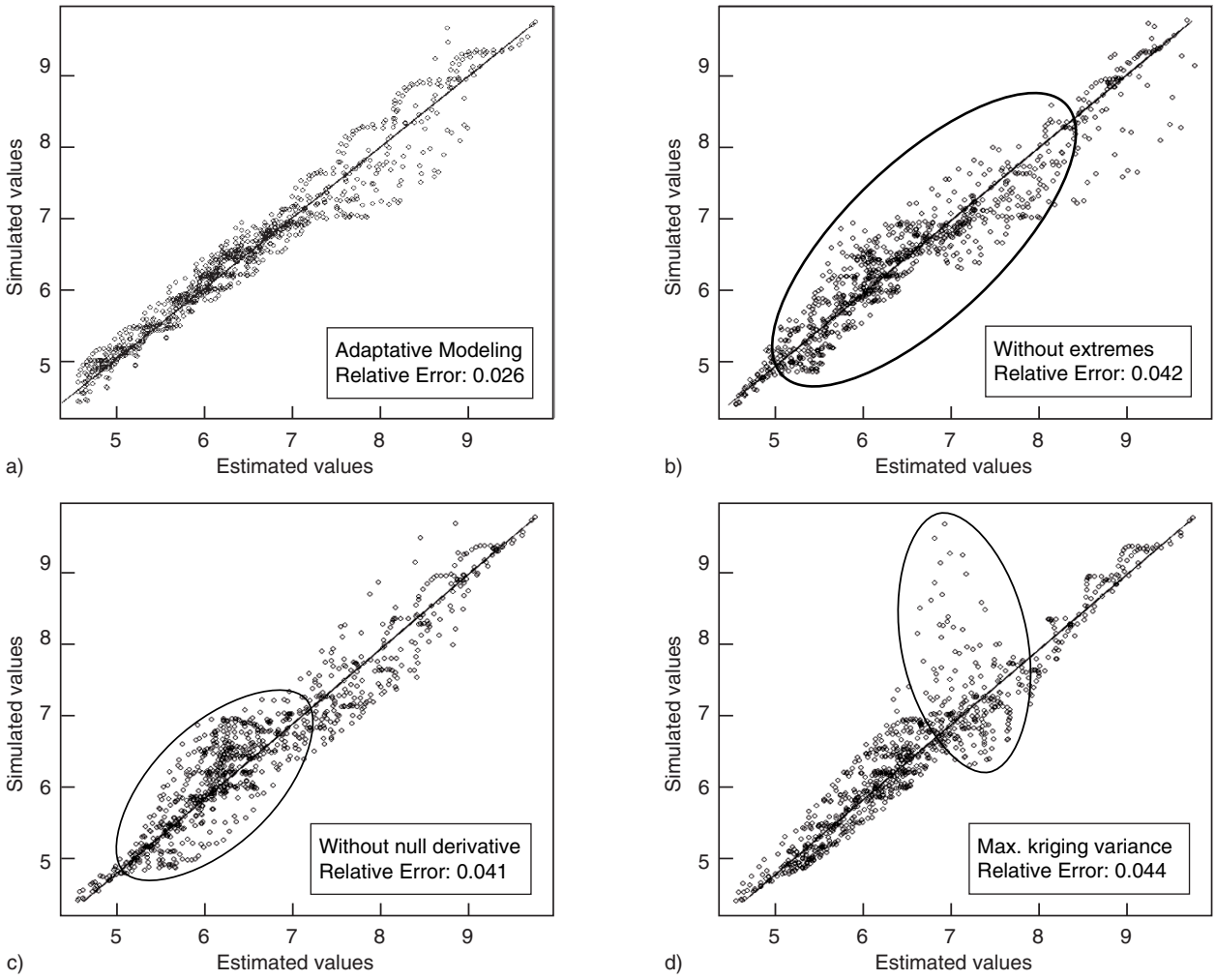

Figure 11

Modeling of the function G: (a) with the adaptive modeling, (b) without addition of extremes, (c) without null derivative points, (d) without testing sensitivity to predictivity with the pilot point method.

quality. Figure 10b shows that the iso-lines of the approximation and the reference are very close to each other, and many irregularities of the response are detected. The global maximum, located in (0.25, 0.6) is also detected through the iso-simplex method, although no point of the initial experimental design was located in this area.

The objective in the next section is to present the impact of each criterion presented in this paper on the accuracy of the approximation (proxy model) $\tilde{f}$.

4.2 Interest of Each Criterion of Adding New Simulations

In this section, we examine the impact of each criterion for adding new simulations. To do so, we model the analytical function G, with 40 simulations using different criteria of adding new simulations. The expression of the function G is:

$$G(x,y) = 0.2e^{(x-3)} + 2.2|y| + 1.3y^6 - 2y^2 - 0.5y^4 - 0.5x^4 + 2.5x^2 + 0.7x^3 + \frac{3}{(8x-2)^2 + (5y-3)^2 + 1} + \sin(5x)\cos(3x^2)$$

The quality of the approximations is presented on Figure 11, using a plot of the predicted values obtained with the approximation versus the simulated values on a grid of 900 points. The first plot in Figure 11a presents the results obtained with the approximation of the function G using the adaptive methodology. Figure 11b shows the approximation fitted using the adaptive methodology but without the simulation of the extremes. Figure 11c presents the results without

incorporating the null derivative points into the approximation. Finally, we simulate directly the point of maximal kriging variance, instead of using pilot points to determine the points most sensitive to perturbations. These results are shown in Figure 11d.

As Figure 11 shows, the results obtained using the adaptive methodology are more accurate than those obtained without using all four criteria. Tests on other analytical functions have resulted in the same conclusion [14]. Each criterion proposed in this paper increases the amount of information brought by the simulations. The extremes and the null derivative points allow a better characterization of the inflexions of the response, whereas the pilot point methodology is crucial in the methodology to explore optimally new areas and to improve the predictivity of the approximation.

4.3 Higher Dimension Experimental Designs – Comparison with Latin Hypercubes

In this section, we compare the results of modelling analytical functions with 2, 3, 4 and 5 parameters using first the adaptive modeling and then Latin hypercubes with universal kriging. The relatives errors obtained with some functions are presented in the Table 1. The number of simulations required is shown in parentheses. The second column represents the results obtained with the adaptive modeling. The third column presents the size required for a Latin hypercube to have an equivalent quality. Finally, the reduction in the number of simulations using the adaptive modeling instead of the Latin hypercube method is given.

TABLE 1

Relative error and # of simulations for many irregular functions

	Adaptive Modeling	Latin Hypercube	Reduction in simulations
2D function #1	0.014 (43)	0.014 (58)	26%
2D function #2	0.016 (58)	0.016 (74)	22%
3D function #1	0.051 (74)	0.051 (120)	38%
3D function #2	0.002 (69)	0.002 (110)	37%
4D function #1	0.162 (100)	0.161 (345)	71%
4D function #2	0.011 (97)	0.019 (405)	76%
5D function	0.016 (134)	0.061 (1350)	90%

Clearly the results obtained using the adaptive method are more accurate than those obtained using Latin hypercubes. Note that because many Latin hypercubes can exist for a given number of simulations and parameters, multiple tests were performed for each analytical function. In Table 1 we show the results for the Latin hypercube with the fewest number of simulations.

Using tests on many analytical functions, we construct a plot representing the number of simulations required as a function of the number of parameters using first the adaptive modeling approach, and then using Latin hypercubes of same quality (relative error).

Figure 12 shows that the number of simulations appears to increase exponentially for the Latin hypercube method and linearly for the adaptive approach. This result suggests that as the number of parameters increases, the greater the benefit is

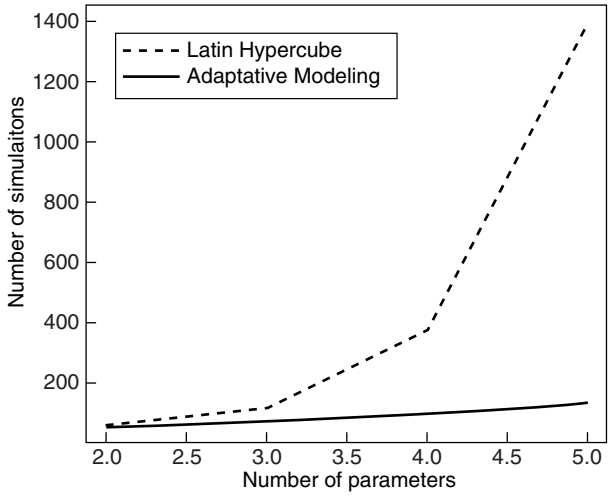

Figure 12

Evolution of the number of simulations as a function of the number of parameters.

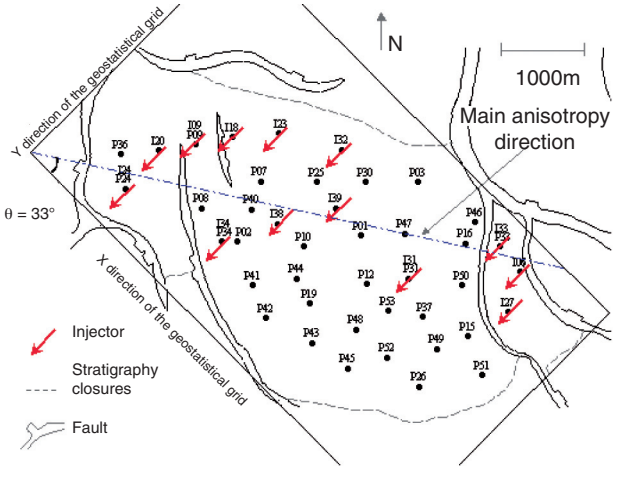

Figure 13

Top view of PBR field.

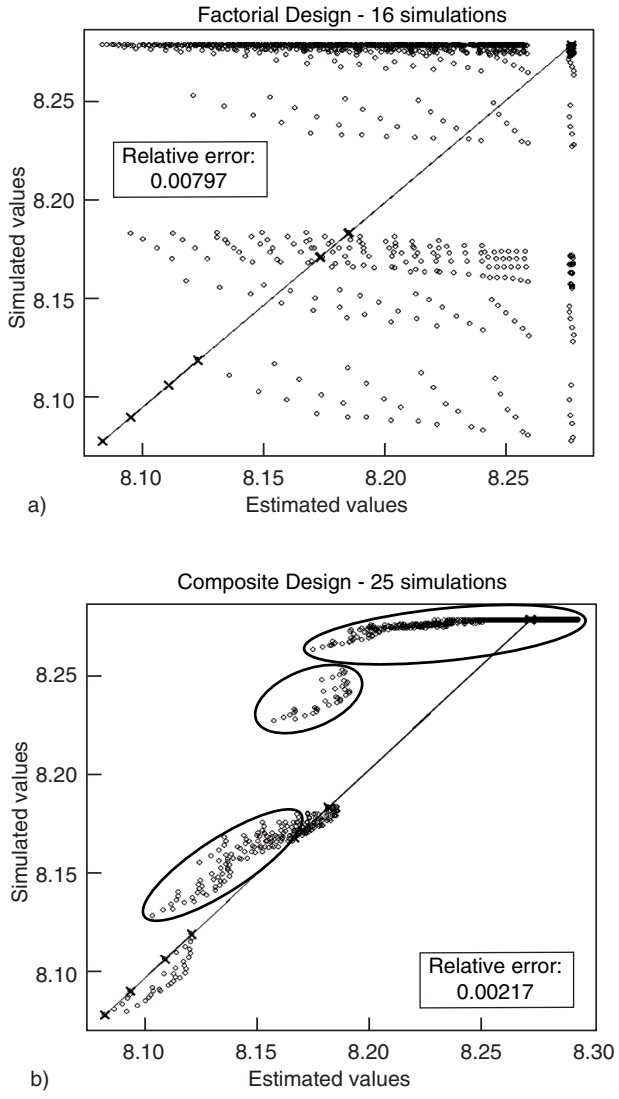

Figure 14

Prediction of CumOil with:
(a) factorial design; (b) composite design.

found using the adaptive approach (in terms of number of flow simulations). In other words, for the same number of simulations performed, the adaptive method provides for a more accurate approximation of the response.

5 NUMERICAL RESULTS FOR A BRAZILIAN RESERVOIR UNCERTAINTY QUANTIFICATION

5.1 Description of the PBR Field

The PBR reservoir is an oil field located offshore Brazil and operated by Petrobras. Figure 13 shows a top view of the PBR field. The reservoir is delineated by structural and stratigraphic traps. The North and South limits are due to stratigraphic closures, whereas the East and West ones are due to fault. The mean depth of the reservoir is about 3100 m and its thickness varies between 90 to 180 m. There are 34 production wells and 13 water injection wells. All the wells are vertical.

For a more detailed description of the PRB field, please refer to Reis *et al*. [15].

5.2 Uncertainty Quantification

The objective of the study is to construct a proxy model of the fluid flow simulator, for a given production response. This model is built in order estimate the risk associated with the uncertainty of the influential fluid flow parameters on the response, which is the cumulative oil production at 6 years.

After a sensitivity study for 18 parameters, 4 parameters were determined to have an important effect on the oil production:

– the water-oil contact depth in the sector 1, named ZOW1;
– the skin factor, named SKIN;
– the maximum relative permeability for gas, named KRGM;
– the maximum relative permeability for water, named KRWM.

The intervals of variation for these parameters is given in Table 2.

TABLE 2

Minimum and maximum value for each uncertain parameter

Parameters	Min	Max
ZOW1 (m)	3 092	3 118
SKIN (adim)	– 5	10
KRGM (adim)	0.4	0.6
KRWM (adim)	0.25	0.45

The objective is therefore to predict, using the adaptive method, an accurate approximation of the cumulative oil production (CumOil) after 6 years as a function of the parameters: ZOW1, SKIN, KRGM, KRWM.

In order to validate the methodology, we performed simulations on a grid $6^4 = 1296$ points. This is usually not possible in reservoir study, because of the time required for each flow simulation. This set of simulations is used as a reference to allow for comparison between different approaches.

We present the results of the proxy models obtained first using the classical approach, which consists of performing simulations following classical experimental designs, then using adaptive modeling, and finally using two types of Latin hypercubes. In all of these cases, the only change concerns the sampling of the experimental domain, the model is obtained using universal kriging *(Equation 1)*.

5.2.1 Classical Experimental Design: Factorial Design and Composite Design

We first perform simulations corresponding to a factorial design with 16 points and a composite design with 25 points. Figure 14 illustrates for both cases the prediction obtained with the proxy model versus the reference case using flow simulations at 1296 points on the grid.

As we can see, both models have poor quality. Prediction of the cumulative oil with classical experimental designs would in this case be unreliable. This is demonstrated in Figure 15. This figure shows the impact of the poor model quality on the probability density estimation. The probability density is calculated using Monte Carlo sampling with uniform distribution between the minimum and maximum values for each parameter. To create the probability density, 10 000 points are selected within the uncertain domain, and the proxy model is evaluated at each point.

The probability density estimated from the proxy model obtained with classical experimental designs is very different from the reference. As shown in Table 3, the estimation of the quantiles P10, P50 and P90 are not reliable - decisions based on this results can be poor.

TABLE 3

Quantiles for classical designs (Mm^3)

	P10	P50	P90
Reference	8.17	8.27	8.28
Factorial	8.14	8.20	8.26
Composite	8.16	8.25	8.27

We now apply the adaptive method in order to improve the quality of the model.

5.2.2 Adaptive Modeling and Latin Hypercube Modeling

For the adaptive model, the initial experimental design is a Latin hypercube with 25 simulations. The design is enhanced at each iteration with 1 maximum and 1 minimum point, 2 points of null partial derivative and the 3 most sensitive points on the predictivity of the approximation, located at the maximum of the kriging variance. The results are presented in Figure 16.

A total of 67 simulations was necessary to obtain an accurate modeling of the cumulative oil production CumOil. The proxy model obtained using the adaptive method is of good quality, with a relative error of 0.00029.

In order to test the performances of the adaptive method, particularly the quality of sampling strategy, we propose to model CumOil using Latin hypercubes requiring the same number of simulations than the adaptive method, *i.e.* 67 simulations. The results are presented in Figure 17.

The approximations fitted according to Latin Hypercubes do not provide a good predictivity of the model, particularly for the minimal values of CumOil. Figure 17 shows that the values below 8.20 Mm^3 are badly estimated, which can have serious implications for decision taking.

After an analysis of the parameters, a large value for the SKIN implies low cumulative oil production. Therefore, it is important to study particularly the risk induced for large values of the SKIN. To do so, we construct a Monte Carlo sampling for which values of the parameters SOW1, KRGM and KRWM are distributed uniformly between the minimum and

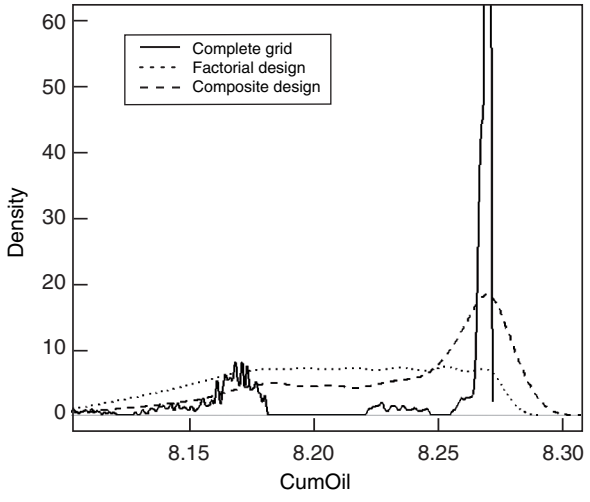

Figure 15

Density probability estimations.

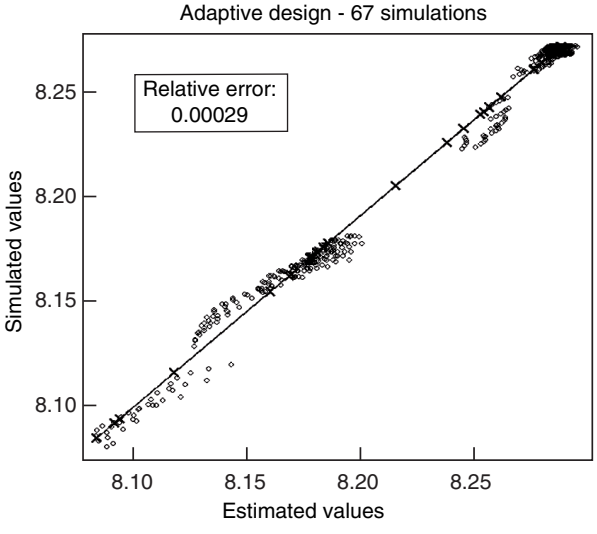

Figure 16

Results of adaptive modeling for Cumoil.

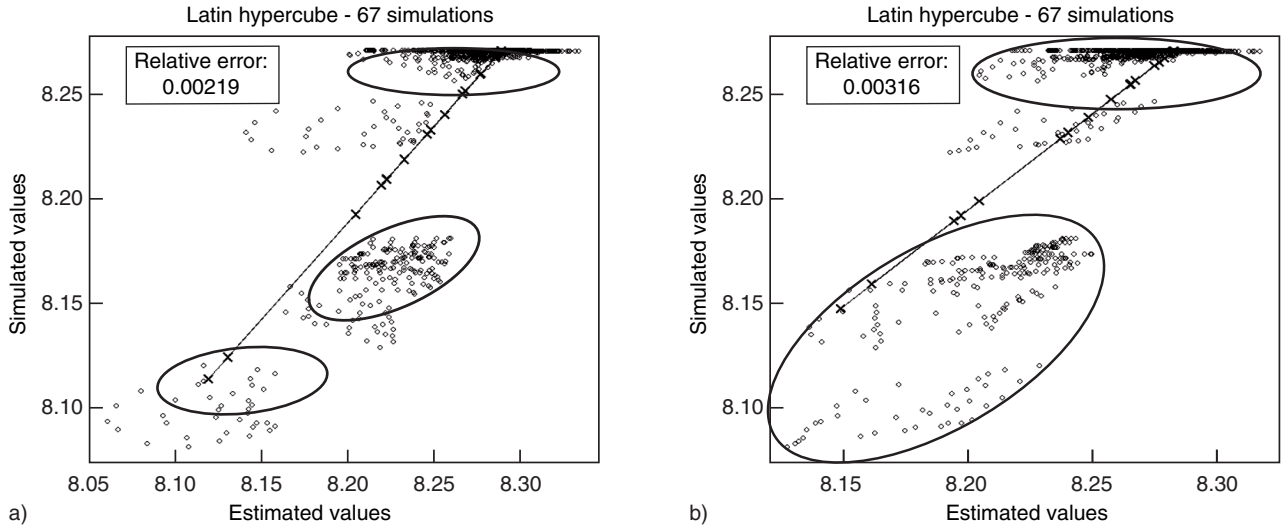

Figure 17

Results of modeling for two Latin hypercubes: (a) Latin Hypercube 1; (b) Latin Hypercube 2.

maximum values in Table 2, and the SKIN value is fixed to its maximum value: 10. As previously, the sample size is 10 000 points. Figure 18 gives the probability density obtained with the two Latin hypercubes, and the results for the adaptive modeling.

Note that in Figure 18 there are significant differences between the probability densities obtained using the two Latin hypercubes and the adaptive modeling. The use of the adaptive modeling captures more accurately the reference probability density. Once more, the quantiles are unreliable, as shown in Table 4. Note that a more accurate probability density results in a more reliable quantification of risk. For example in this case, with an oil barrel at $ 70, an overestimation of 50 000 m^3 in production forecasts leads to an overestimation of $ 22 015 000, which may considerably change decisions.

TABLE 4

Quantiles for adaptive design and Latin hypercubes (Mm^3)

	P10	P50	P90
Reference	8.101	8.164	8.176
HL1	8.166	8.207	8.289
HL2	8.163	8.211	8.229
Adaptive	8.120	8.165	8.179

Note that Latin hypercubes are generated using a stochastic algorithm using a random number generator. For a different initial seed, the Latin hypercube samples the uncertainty space differently. In addition, the correct number of simulations is difficult to know *a priori*, and may be computationally time-prohibitive. Thus, using uniquely Latin Hypercubes for uncertainty quantification may not give reliable and repeatable results.

6. DISCUSSION

In this paper we have presented a methodology for uncertainty quantification and risk assessment through the construction of a reliable proxy model of the fluid flow simulator. This proxy can be considered as a good representation of

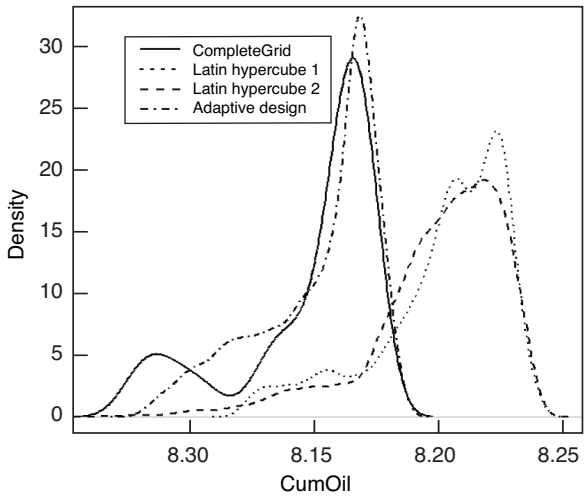

Figure 18

Density probability estimations.

the fluid flow simulator, whose evaluation is not expensive and therefore allows for risk analysis using Monte Carlo sampling. In addition to risk analysis, this method can also be used to optimize production schemes. For example, one can determine the value of the control parameter (*e.g.* location of infill well) in order to maximize the production while minimizing the risk associated with uncertainty. In doing so, the impact of uncertainty on production forecast is reduced and decisions are made that enable a compromise between maximum production and minimal risk.

In Section 3.2 above, we have shown how the pilot point method can be used to test for the sensitivity of the pilot point location and value on the prior predictivity of the model approximation. In addition to testing the sensitivity, the pilot point methodology can be use to optimize the prior predictivity of the proxy model. By varying the fictitious value at the pilot point, a standard optimization procedure can be employed to maximize the common predictivity coefficient Q_2. By doing so, this procedure ensures a better robustness of the approximation without adding additional flow simulations. Scheidt [14] provides further detail on the pilot point optimization, and examples of its application.

Future work includes researching methods to systematically determine sources of irregularities. The adaptive method could improve its effectiveness by evaluating first the regularity or the irregularity of the response for each parameter, using a sensitivity analysis (Sobol', Fast) for example. This would allow a more intensive sampling in the subdomain defined by the most irregular parameters. An alternative could be to reduce the dimension of the uncertain domain using principle axes defined by screening methods such as canonical analysis or principal component analysis. This should allow the addition of new simulations in the subdomain, and therefore to reduce the total number of simulations. See [12] for preliminary work in this area.

Other improvement of the methodology consists in taking advantage of different kriging operators proposed in the bibliography. For instance, the use of kriging with moving neighborhood in the over-sampled areas allows for local trend fitting. However, it should be noted that in the case where data is sparse, universal kriging should be preferred.

Finally, the sampling strategy proposed in this paper is designed to optimally capture the behavior of an unique response. Therefore, this sampling may not be optimal to catch the complexity of other production responses. To get a global overview of the risk, it would be of interest to be able to consider multiple responses.

CONCLUSION

Modeling of irregular production response surfaces is a complex problem due to practical limits on the number of fluid flow simulations that can be performed. The common approach consists in building proxy models using regression or kriging. These models are created with a limited number of simulations defined by classical experimental design or quasi Monte Carlo sampling with Latin hypercubes. A major drawback of these approaches is that they require a prior knowledge of the dependence between the parameters and the response, and find the limits of practicality when the response is complex.

This work presents a new approach for uncertainty quantification to properly assess risk even if the impact of uncertain parameters is complex and the response is irregular. The adaptive modeling method combines the advantages of geostatistical interpolation techniques and experimental designs. Since the response is deterministic (*i.e.* the fluid flow simulator gives a unique response given an identical set of parameters), universal kriging is used to model the response. The requirement of a minimal amount of flow simulations to model the response leads to using an experimental design approach. The use of evolutive designs maximizes the amount of information obtained for a given simulation. In addition, the adaptive aspect of the methodology allows the control of the number of simulations. The reservoir engineer can decide either to add simulations to get a more accurate approximation, or to sacrifice the approximation quality in order to reduce the overall simulation time.

The method was validated on many analytic functions and with a real field case from offshore Brazil. The method enabled an efficient modeling of highly irregular responses, with a better approximation of the response compared to Latin hypercube of the same size. In addition, results show, for same quality of predictivity, a dramatic decrease in the number of simulations compared to classical designs. This work has the potential to improve considerably the effectiveness of decision making while accounting for uncertainty.

ACKNOWLEDGEMENTS

The authors would like to thank Petrobras for the publication of the results of the PBR case. This work was supported by the partner of the COUGAR II JIP: *BHP-Billiton, ConocoPhillips, Gaz de France, Petrobras, Repsol-Ypf* and *Saudi Aramco*. Part of this work has been performed thanks to the use of the COUGAR™ software for risk analysis in reservoir engineering, software developed by IFP and currently commercialized by *Schlumberger SIS*.

REFERENCES

1 Manceau, E., Zabalza-Mezghani, I., and Feraille M. (2002) *Cougar_Opt software user Manual,* Version 2.0, IFP.

2 Manceau, E., Feraille, M., Zabalza-Mezghani, I., Portella, R. and Reis, L.C. (2005) *Advanced risk analysis approach for optimization of a water injection program - Illustration on a Brazilian field case*, SPE-94845.

3 Johnson, M. E., Moore, L. M. and Ylvisaker, D. (1990) Minimax and Maximin distance designs. *J. Stat. Plan. Infer.,* **26**, 131-148.

4 Sacks, J., Schiller, S. and Welch, W. (1989) Design for computer Experiments. *Technometrics*, **31**, 41-47.

5 Sacks, J., Welch, W., Mitchell, T., and Wynn, P. (1989) Design and Analysis of Computer Experiments (with discussion). *Stat. Sci.,* **4**, 409-435.

6 Chilès J.-P. and Delfiner, P. (1999) *Geostatistics, Modeling Spatial Uncertainty,* Wiley Series in probability and statistics.

7 Nelder, J.A., Mead, R. (1965) A simplex Method for Optimization. *Comput. J.,* **7**, 308.

8 Currin, C., Mitchell, T.J., Morris, M.D. and Ylvisaker, D. (1991) Bayesian prediction of deterministic functions, with an application to the design an analysis of computer experiments. *J. Am. Stat. Assoc.,* **86**, 953-963.

9 Fang, K.T., Li, R. and Sudjiants, A. (2005) *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC; London.

10 Santner, T.J., Williams, B.J. and Notz, W.I. (2003) *The Design and Analysis of Computer Experiments,* Springer Series in Statistics.

11 RamaRao, B.-S., La Venue, A.-M., De Marsily, G. and Marietta, M.-G., (1995) Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. *Water Resour. Res.*, **31**, 475-493.

12 Montgomery, D.C. and Peck, E.A. (1991) *Introduction to linear regression analysis*, second edition, J. Wiley.

13 Scheidt C., and Zabalza-Mezghani I. (2004) *Assessing Uncertainty and Optimizing Production schemes - Experimental Designs for non-linear production response modeling - An application to early water breakthrough prevention*, ECMOR IX, Cannes, 30 August-2 September 2004, France.

14 Scheidt, C. (2006) *Analyse Statistique d'expériences simulées : Modélisation adaptative de réponses non-régulières par krigeage et plans d'expériences – Application à la quantification des incertitudes en ingénierie des réservoirs pétroliers,* Thesis IFP - Université Louis Pasteur de Strasbourg.

15 Reis, L.C., Hu, L.Y., de Marsily, G. and Eschard, R. (2000) *Production Data Integration Using a Gradual Deformation Approach: Application to an Oil Field (Offshore Brazil)*, SPE 63064.

Final manuscript received in November 2006

Copyright © 2007 Institut français du pétrole

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than IFP must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee: Request permission from Documentation, Institut français du pétrole, fax. +33 1 47 52 70 78, or revueogst@ifp.fr.

APPENDIX A: VARIOGRAM FITTING

One of the difficulties in using kriging is the estimation of the variogram. We propose two different ways for building a variogram.

Case 1: Few simulations in the current experimental design $X_{(i)}$

Usually, at the beginning of the iterative process, the current experimental design contains few simulations and thus it is not possible to estimate empirically the covariance function. In this case, we use the approach described in [10]. These authors suggests assuming that the covariance is Gaussian with the form:

$$Corr(\boldsymbol{h}) = \sigma^2 \exp\left\{-\sum_{i=1}^{k}\left(\frac{h_l}{\theta_l}\right)^2\right\} \text{ with } h_i = \left|x_{jl} - x_{kl}\right| \quad \forall k, j \le n$$

In addition, we propose to estimate the value of the parameters θ_i by the restricted maximum likelihood method (Patterson and Thompson in 1971). It has also been called the marginal maximum likelihood estimate [10].

An estimate of the variance of the process is given by :

$$\tilde{\sigma}^2_{MLE} = \tilde{\sigma}^2_{MLE}(\theta) = \frac{1}{n-q}\,{}^t(\boldsymbol{f} - \boldsymbol{G}\hat{\beta})\boldsymbol{R}^{-1}(\boldsymbol{f} - \boldsymbol{G}\hat{\beta})$$

where q = range($\boldsymbol{G}$) and n the total number for simulations.

Using this expression, it is straightforward to deduce the value of $\tilde{\theta}_{MLE}$ by the minimizing the function:

$$\theta_{MLE} = \min_{\theta}\left\{(n-q)\log\left(\tilde{\sigma}^2_{MLE}(\theta)\right) + \log\left(\det\left(\boldsymbol{R}(\theta)\right)\right)\right\}$$

Case 2: Large number of simulations in the current experimental design $X_{(i)}$

After several iterations, the experimental design may have a sufficient number of simulations to challenge the assumption that the covariance is Gaussian. In this case, we propose to determine the covariance function which minimizes an objective function J that represents the squared sum of the error between the empirical variogram and the theoretical variogram:

$$J = \sum_{l}\left(\gamma(\boldsymbol{h}^l, \theta) - \gamma_e(\boldsymbol{h}^l)\right)^2$$

where:

- $\gamma_e(\boldsymbol{h}^l)$ represents the empirical variogram obtained according to the simulated runs,
- $\gamma_e(\boldsymbol{h}^l, \theta)$ is a theoretical variogram usually used in reservoir engineering: exponential, Gaussian or spherical,
- $\boldsymbol{h}^l$ represents the distances used for the construction of the empirical variogram.

However, using a single theoretical variogram is not always optimal *(Fig. 19a)*. Therefore, we propose to generalize the objective function J with the estimation of a linear combination of these theoretical variogram expressions. The variogram to optimize is then:

$$\gamma(\boldsymbol{h}, \theta, \boldsymbol{c}) = \lambda_1\gamma_{exp}\left(\boldsymbol{h}, \theta_1, c_1, p\right) + \lambda_2\gamma_{sph}\left(\boldsymbol{h}, \theta_2, c_2\right)$$

where:

$$\gamma_{exp}\left(\boldsymbol{h}, \theta, c, p\right) = c\left(1 - \exp\left\{-\left(\frac{\boldsymbol{h}}{\theta}\right)^p\right\}\right) \text{ and } 0 < p \le 2.$$

The objective is to evaluate the values of the parameters λ_1, λ_2, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{c}$ and p in order to minimize the function:

$$(\theta, \boldsymbol{c}, p) = \min\left\{\sum_i\left(\gamma(\boldsymbol{h}^i, \theta, \boldsymbol{c}, p) - \gamma_e(\boldsymbol{h}^i)\right)^2\right\}$$

As we can see in Figure 19b, the use of this kind of variogram allows a better fit of the experimental data - the error is reduced. Therefore, when the number of simulations of the experimental design $X_{(i)}$ is sufficient, we perform an optimization of the variogram, avoiding a prior assumption on the choice of a particular variogram model.

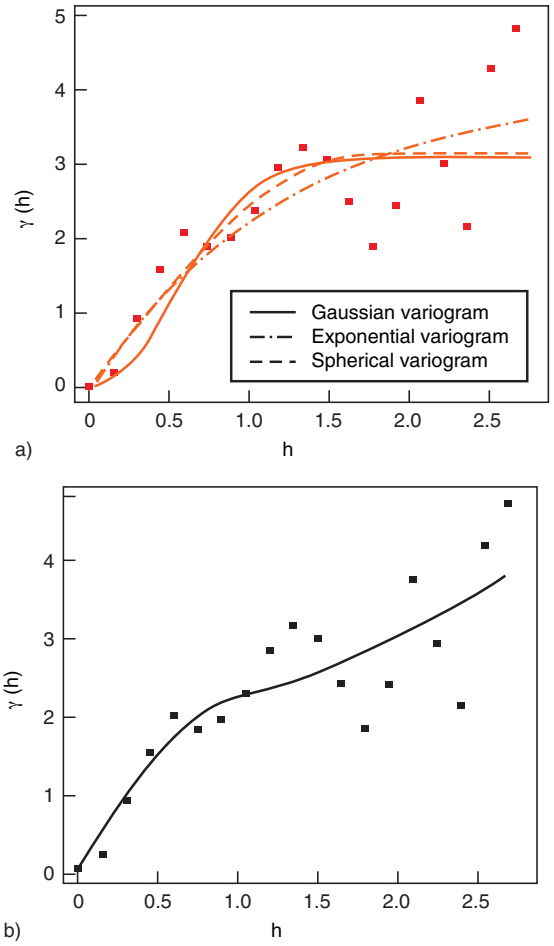

Figure 19

Variogram fitting: (a) Gaussian, exponential and spherical theoretical variogram; (b) linear combination of theoretical variograms.