Uncertainty Evaluation in Reservoir Forecasting by Bayes Linear Methodology

Daniel Busby[1], Chris L. Farmer[1,2], and Armin Iske[3]

[1] Schlumberger Abingdon Technology Center, Abingdon OX14 1UJ, UK,
  {dbusby4,farmer5}@slb.com
[2] Oxford Centre for Industrial and Applied Mathematics, University of Oxford,
  Oxford OX1 3LB, UK, farmer5@slb.com
[3] Department of Mathematics, University of Hamburg, D-20146 Hamburg
  iske@math.uni-hamburg.de

Summary. We propose application of Bayes linear methodology to uncertainty evaluation in reservoir forecasting. On the basis of this statistical model, effective emulators are constructed. The resulting statistical method is illustrated by application to a commonly used test case scenario, called PUNQS [11]. A statistical data analysis of different output responses is performed. Responses obtained from our emulator are compared with both true responses and with responses obtained using the response surface methodology (RSM), the basic method used by leading commercial software packages.

1 Introduction

A reservoir simulator is a large computer code which requires solving a system of nonlinear partial differential equations from complex geological model data. The reservoir geology is typically characterized by a huge number of input parameters to the simulator. As these input parameters are usually uncertain, so is the output of the simulator uncertain. Thus, uncertainty evaluation of large simulation codes has become a major task in reservoir forecasting.

In this paper Bayes linear methodology is applied to reservoir forecasting using a sequential experimental design [9] for the construction of effective emulators. We remark that the application of the Bayes linear approach to comparable applications was recently discussed in related works [3, 7]. Moreover, our sequential experimental design is similar to that one in [13].

The performance of Bayes linear methodology is evaluated by comparison with true responses for different outputs of the reservoir simulator. Moreover, response surfaces from reservoir forecasting are analyzed, and our results are also compared with the *response surface methodology* (RSM) [6], which is the basic method of the commercial software package COUGAR [2].

The outline of this paper is as follows. In Section 2, the methodology of Bayes linear estimation is reviewed. In Section 3, a model for the construction of effective emulators, based on the Bayes linear estimator, is proposed. Numerical results are in Section 4, where numerical comparisons with the response surface methodology (RSM) are performed.

2 Bayes Linear Methodology

Simulator output $s(\mathbf{x})$ is a function of n, $n \geq 1$, uncertain input parameters $\mathbf{x} \in \chi \subset \mathbb{R}^n$. Uncertainty evaluation requires the *probability density*

$$p(y) = p(s(\mathbf{x}) = y) = \int_\chi \delta(s(\mathbf{x}) - y)\rho(\mathbf{x})d\mathbf{x},$$

where $\rho(\mathbf{x})$ is a given density function of $\mathbf{x} \in \chi$ and δ is the Dirac δ-functional. Statistical quantities, such as expectation, $\mathrm{E}[s(\mathbf{x})]$, or variance, $\mathrm{Var}[s(\mathbf{x})]$, are also of particular interest,

$$\mathrm{E}[s(\mathbf{x})] = \int_\chi s(\mathbf{x})\rho(\mathbf{x})d\mathbf{x},$$

$$\mathrm{Var}[s(\mathbf{x})] = \int_\chi \left|s(\mathbf{x}) - \mathrm{E}[s(\mathbf{x})]\right|^2 \rho(\mathbf{x})d\mathbf{x}.$$

For these tasks, Monte Carlo methods are computationally too expensive, as too many simulation runs are required. As shown in [3, 6, 8], more sophisticated statistical approaches, such as response surface methodology (RSM) or Bayesian approaches, are more appropriate than Monte Carlo methods.

When $s(\mathbf{x})$ is a smooth function, one can use multiple regression techniques to approximate $s(\mathbf{x})$ from a few simulation runs. In the RSM, a linear model is used, i.e., a linear combination of q fixed basis functions; usually low order polynomials. The coefficients of the linear model are calculated using a standard least squares technique.

RSM was originally introduced in physical experiments, where each observation of a physical process is subject to measurement error. In contrast, a simulator is deterministic, i.e., rerunning the code with the same inputs gives identical observations. In this case, an interpolatory estimator rather than an approximation is usually preferred. A Bayesian approach yields, unlike RSM, an *interpolatory* (posterior) estimator, see the appendix of [3] for details.

Application of a Bayesian approach results in updating a prior distribution of a statistical model s_B by *Bayes' rule*,

$$\mathrm{P}_{\mathrm{Post}}(s_B(\mathbf{x})|s_{\mathbf{X}}) \propto \mathrm{P}_{\mathrm{Prior}}(s_B(\mathbf{x})) \, \mathrm{P}_{\mathrm{Likelihood}}(s_{\mathbf{X}}|s_B(\mathbf{x})),$$

where $s_{\mathbf{X}} = (s(\mathbf{x}_1), \ldots, s(\mathbf{x}_m))^T \in \mathbb{R}^m$ denotes a response vector containing m simulation outputs taken at a design set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subset \mathbb{R}^n$ of m pairwise distinct input configurations, and P is the (conditional) probability.

We prefer to work with a Bayes *linear* estimator, as suggested in [3]. This is mainly for computational reasons, as the Bayes linear estimator s_{BL} considers *only* the first two moments of the prior and posterior distribution, which are related by

$$\mathrm{E}[s_{\mathrm{BL}}(\mathbf{x})|s_{\mathbf{X}}] = \mathrm{E}[s_{\mathrm{BL}}(\mathbf{x})] + \mathrm{Cov}[s_{\mathrm{BL}}(\mathbf{x}), s_{\mathbf{X}}]\mathrm{Var}[s_{\mathbf{X}}]^{-1}(s_{\mathbf{X}} - \mathrm{E}[s_{\mathbf{X}}]),$$
$$\mathrm{Var}[s_{\mathrm{BL}}(\mathbf{x})|s_{\mathbf{X}}] = \mathrm{Var}[s_{\mathrm{BL}}(\mathbf{x})] + \mathrm{Cov}[s_{\mathrm{BL}}(\mathbf{x}), s_{\mathbf{X}}]\mathrm{Var}[s_{\mathbf{X}}]^{-1}\mathrm{Cov}[s_{\mathbf{X}}, s_{\mathrm{BL}}(\mathbf{x})].$$

Therefore, Bayes linear estimation can be viewed as an approximation to a full Bayesian approach. Moreover, we remark that in the absence of any prior information on model parameters for mean and autocovariance, the Bayes linear methodology is equivalent to (universal) kriging, see [5] for details.

Now the random process $s_{BL}(\mathbf{x})$ with posterior mean $\mathrm{E}[s_{BL}(\mathbf{x})|s_{\mathbf{X}}]$ and variance $\mathrm{Var}[s_{BL}(\mathbf{x})|s_{\mathbf{X}}]$ is referred to as an *emulator*. An emulator is a cheap surrogate for a (costly) simulator.

3 Construction of the Emulator

3.1 Model Description

Similarly to [3], we work with a (prior) emulator of the form

$$s_{\mathrm{BL}}(\mathbf{x}) = \beta^T g(\mathbf{x}_*) + \epsilon(\mathbf{x}_*), \qquad (1)$$

with unknown coefficients $\beta \in \mathbb{R}^q$, $q < m$, regression functions $g = (g_1, \ldots, g_q)$, and where $\mathbf{x}_*$ are the active variables of $\mathbf{x} \in \chi$. Loosely speaking, the active variables are those which account for most of the output variation. The discrepancy between the linear regression $\beta^T g(\mathbf{x}_*)$ and the simulator $s(\mathbf{x})$ is modelled by a stationary Gaussian process $\epsilon(\mathbf{x}_*)$ with zero mean and an auto-covariance function

$$\mathrm{Cov}[\epsilon(\mathbf{x}_*), \epsilon(\mathbf{y}_*)] = \sigma_\epsilon^2 r(\mathbf{x}_* - \mathbf{y}_*),$$

where $r(\mathbf{z})$ denotes a correlation function to be specified. The selection of active variables $\mathbf{x}_*$, of the regression functions g and of the correlation function $r(\mathbf{z})$ are based on prior knowledge about the process. This is discussed in the following subsection.

3.2 The Prior Summaries

Prior knowledge about the random process is usually built by expert elicitation [4]. In our case, an initial set of simulator runs is used to support the elicitation process. This initial data is not analyzed statistically. The data is rather interpreted by reservoir engineers who provide estimates of the prior

mean $\mathrm{E}[s_{\mathrm{BL}}(\mathbf{x})]$ and variance $\mathrm{Var}[s_{\mathrm{BL}}(\mathbf{x})]$. The required selection of the active variables $\mathbf{x}_*$ and of the regression functions g in (1), usually low order polynomials, is done through sensitivity analysis, as described in [10, 14].

We decided to work with the autocovariance function

$$\mathrm{Cov}[\epsilon(\mathbf{x}_*), \epsilon(\mathbf{y}_*)] = \sigma_\epsilon^2 \exp(-\theta\|\mathbf{x}_* - \mathbf{y}_*\|), \tag{2}$$

which leads to continuous but non-smooth response surfaces, as desired in the situation of our particular application, see Section 4. In a more general situation, the selection of the autocovariance function in (2) should be made on the basis of previous observations in similar problems.

The parameters θ and σ_ϵ in (2) can be determined by *maximum likelihood estimation* (MLE), see [13]. This gives

$$\hat{\sigma}_\epsilon^2 = \frac{1}{m}(s_{\mathbf{X}} - G\beta)^T R^{-1}(s_{\mathbf{X}} - G\beta),$$

for the estimation of σ_ϵ^2, where m is the number of simulations, and where

$$R = (r(\mathbf{x}_i, \mathbf{x}_j))_{1 \le i,j \le m} \in \mathbb{R}^{m \times m}, \qquad G = (g_j(\mathbf{x}_i))_{1 \le i \le m; 1 \le j \le q} \in \mathbb{R}^{m \times q}.$$

Estimation of θ by MLE requires global optimization and is generally sensitive to the number of simulations. Therefore, in our case we prefer to use data visualization techniques which yields a more robust estimate of $\hat{\theta} = 2$ for θ. For more details on the estimation of the autocovariance function in (2) we refer to our previous paper [1].

3.3 Experimental Design

In computer simulations, the goal of experimental design is to determine suitable input configurations for effective data analysis. The required data analysis is specific to the objectives of the experiment. Possible objectives include uncertainty propagation, optimization of certain response functionals (e.g. oil production), and tuning the simulator to physical data, *history matching.*

In reservoir forecasting, experimental design is of primary importance, especially since each simulation run is computationally very expensive. In view of uncertainty evaluation, we are aiming at the construction of a sufficiently accurate emulator to predict responses at untried input. But we wish to keep the number of required simulation runs as small as possible.

Possible experimental designs can be split in two different categories: single stage methods, such as fractional factorial designs (FFD) or Latin hypercube designs (LHC), and sequential designs which aim at minimizing uncertainty measures of the emulator. In the approach proposed in this paper, a number of initial simulator runs are first performed by using FFD. Then, a number of subsequent simulator runs are done by using a sequential design. But this requires a specific design criterion.

The design criterion we work with relies on the *maximum mean square error* (MMSE). In this case, design points, $\mathbf{x}^*$, are sequentially added, one at a time, where the posterior variance $\mathrm{Var}[s_{\mathrm{BL}}(\mathbf{x})|s_{\mathbf{X}}]$ of the current Bayes linear emulator $s_{\mathrm{BL}} \equiv s_{\mathrm{BL}}^{(m)}$ is maximal among all $\mathbf{x} \in \chi$. In this way, the prediction error of the subsequent (posterior) emulator $s_{\mathrm{BL}}^{(m+1)}$ vanishes at $\mathbf{x}^*$. A similar design criterion is proposed in [13], but for kriging.

In summary, each step of the sequential design is performed as follows.

(1) Compute an input configuration $\mathbf{x}^*$ which maximizes $\mathrm{Var}[s_{\mathrm{BL}}(\mathbf{x})|s_{\mathbf{X}}]$;
(2) Run the simulator at the selected configuration $\mathbf{x}^*$ to obtain $s(\mathbf{x}^*)$;
(3) Rebuild the emulator by including the new simulator output $s(\mathbf{x}^*)$.

As regards a stopping criterion, we chose a customized diagnostic measure which relies on the prediction error

$$\eta(m) = |s_{\mathrm{BL}}^{(m-1)}(\mathbf{x}_m) - s(\mathbf{x}_m)|,$$

where $\mathbf{x}_m = \mathbf{x}^*$ denotes the design point which was added at step m, and $s(\mathbf{x}_m)$ is the simulator response at $\mathbf{x}_m$. Note that $s_{\mathrm{BL}}^{(m)}(\mathbf{x}_m) = s(\mathbf{x}_m)$. When the sequence $\eta(m)$ of prediction errors *stabilizes*, i.e., $|\eta(m) - \eta(m-1)| < \mathrm{TOL}$ for some tolerance TOL, we take $s_{\mathrm{BL}}^{(m)}$ as an a sufficiently accurate emulator.

4 Numerical Results for the PUNQS Test Case

4.1 Reservoir Model Description

The PUNQS test case relies on a synthetic reservoir model taken from the North Sea Brent reservoir, a real-world oilfield. The PUNQS test case is frequently used as an industrial reservoir engineering model since its use in the European research project PUNQ [11] is a benchmark test for comparative inversion studies and for stochastic reservoir modelling.

A top structure map of the PUNQS reservoir field is shown in Figure 1. The geological model contains $19 \times 28 \times 5 = 2660$ grid blocks, 1761 of which are active. The reservoir is surrounded by a strong aquifer in the North and in the West, and it is bounded by a fault to the East and to the South. A small gas cap is located in the centre of this dome-shaped structure. The geological model consists of five independent layers, where the porosity distribution in each layer was modelled by geostatistical simulation. Initially, the field contains six production wells located around the gas-oil contact. Due to the strong aquifer, no injection wells are required.

As suggested by reservoir engineers, we consider the following seven main sources of uncertainty: (i) the analytical coefficient of the aquifer strength, **AQU**, (ii) the residual gas oil saturation, **GOS**, (iii) the residual water oil saturation, **WOS**, (iv) the vertical permeability multiplier in low quality sands, **VPML**, (v) the vertical permeability multiplier in high quality
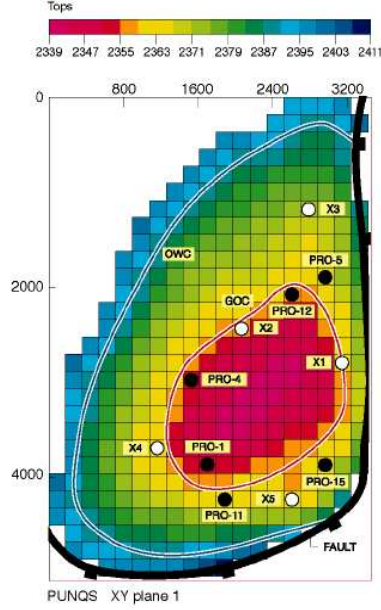
Fig. 1. PUNQS test case. Top structure map of the reservoir field.

sands, **VPMH**, (vi) the horizontal permeability multiplier in low quality sands, **HPML**, (vii) the horizontal permeability multiplier in high quality sands, **HPMH**. For each of the seven input variables, a uniform distribution in the parameter interval $[-1, 1]$ is assumed.

To evaluate and compare different methods by their emulator accuracy, we decided to work with three different error measures when recording the resulting prediction errors for an emulator s_{E}. The error measures are the *mean absolute error*

$$\eta_1 = \|s - s_{\mathrm{E}}\|_1 / |\Xi| = \frac{1}{|\Xi|} \sum_{\mathbf{x} \in \Xi} |s(\mathbf{x}) - s_{\mathrm{E}}(\mathbf{x})|,$$

mean square error,

$$\eta_2^2 = \|s - s_{\mathrm{E}}\|_2^2 / |\Xi| = \frac{1}{|\Xi|} \sum_{\mathbf{x} \in \Xi} |s(\mathbf{x}) - s_{\mathrm{E}}(\mathbf{x})|^2,$$

and *maximum error*,

$$\eta_\infty = \|s - s_{\mathrm{E}}\|_\infty = \max_{\mathbf{x} \in \Xi} |s(\mathbf{x}) - s_{\mathrm{E}}(\mathbf{x})|,$$

where Ξ denotes a fine uniform grid contained in the computational domain χ. We have implemented the proposed approach in the language R [12].

4.2 Numerical Results from Two-Dimensional Input

In this subsection, we present numerical results for two different responses in the PUNQS model from 2D input. The small size of the PUNQS reservoir model, containing only less than $20,000$ grid cells, allows us to perform several thousand simulation runs, which are included in the two numerical tests. The responses from these simulations are taken to visualize the *real* response surface, whose graph is then compared with both the graph of the Bayes linear emulator s_{BL} and the graph of the emulator s_{RSM} obtained by the response surface methodology (RSM).

To demonstrate the good performance of the proposed Bayes linear approach, we selected two rather challenging test cases involving rough response surfaces $s(\mathbf{x})$ of high variation.

The first test case is concerning the oil production rate at well **PRO15** (see Figure 1 bottom right) after 13 years, response surface **P15OPR**, as a function of its two main active variables, **HPMH** and **HPML**. The design set X was constructed by applying FFD to obtain an initial set of 7 points, followed by a sequential design for further 5 points, yielding $m = 12$ design points in total.

Figure 2 displays the response surface of the Bayes linear emulator, s_{BL}, and the response surface obtained by RSM, emulator s_{RSM}. For comparison, Figure 2 displays 10×10 grid points of the true response surface.

| (a) | (b) |
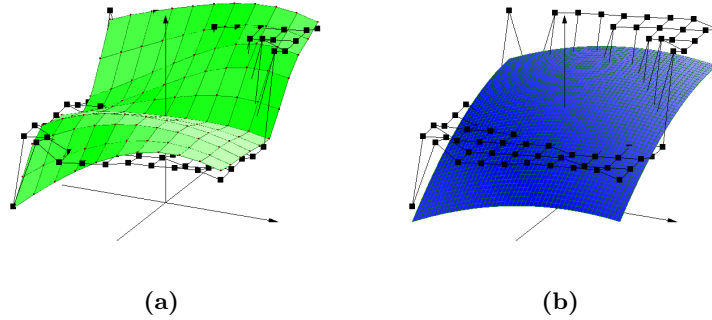
Fig. 2. PUNQS test case **P15OPR(HPMH,HPML).** Response surface of **(a)** Bayes linear emulator s_{BL}, **(b)** s_{RSM}, each constructed by using 12 design points. A 10×10 mesh grid of the true response surface **P15OPR** is shown for comparison.

Note that the response surface s_{BL} obtained from the Bayes linear estimator (Figure 2 **(a)**) is, in comparison with s_{RSM} of RSM (Figure 2 **(b)**), much closer to the true response surface **P15OPR**, and so the Bayes linear esti-

mator is superior. This is also confirmed by our numerical results in Table 1, where their prediction errors η_1, η_2, and η_∞ are shown.

Table 1. PUNQS test case **P15OPR(HPMH,HPML).** Prediction errors from emulators s_{BL} and s_{RSM}, each constructed by using $m = 12$ design points.

Method	η_1	η_2	η_∞
BL	3.0	4.6	17.6
RSM	6.2	7.1	16.2

In our second test case, we consider the bottom hole pressure at well **PRO15** after 13 years, response surface **P15BHP**, as a function of **HPMH** and **GOS**. The design set X was constructed by applying FFD to obtain an initial set of 7 points, followed by a sequential design for further 2 points, yielding $m = 9$ design points in total.

Figure 3 displays the response surface of the Bayes linear emulator, s_{BL}, and the response surface obtained by RSM, emulator s_{RSM}, each of which was constructed by using $m = 9$ design points. For comparison, Figure 3 displays 9×9 grid points of the true response surface. Our numerical results are shown in Table 2.

(a)   (b)
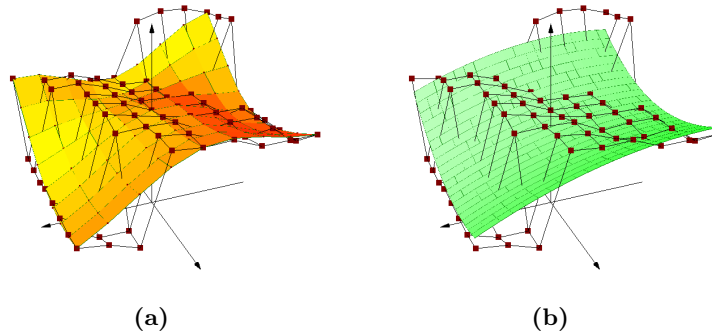
Fig. 3. PUNQS test case **P15BHP(HPMH,GOS).** Response surface of **(a)** Bayes linear emulator s_{BL}, **(b)** s_{RSM}, each constructed by using $m = 9$ design points. A 9×9 mesh grid of the true response surface **P15OPR** is shown for comparison.

Note that the Bayes linear estimator continues to be superior to RSM in terms of its better reconstruction quality. This is supported by both the response surface graphs in Figure 3 and the numerical results in Table 2. Table 2 shows the prediction errors η_1, η_2 and η_∞ obtained from the two

Table 2. PUNQS test case **P15BHP(HPMH,GOS).** Prediction errors from emulators s_{BL} and s_{RSM}, constructed by using $m = 7$ and $m = 9$ design points each.

Method	m	η_1	η_2	η_∞	m	η_1	η_2	η_∞
BL	7	3.7	5.3	13.4	9	2.7	4.3	12.6
RSM	7	4.2	5.5	12.4	9	3.6	4.8	11.3

different emulators, s_{BL} and s_{RSM}. Note that Table 2 involves two different comparisons, one using the initial set of $m = 7$ design points, the other using all $m = 9$ design points. Note that the accuracy of the emulator s_{BL} is, unlike that of s_{RSM}, significantly improved by the adaptive insertion of only two design points, $\mathbf{x}_8$ and $\mathbf{x}_9$. Moreover, the prediction quality of the Bayes linear emulator s_{BL} is superior to that of s_{RSM} not only in smooth regions of the true surface **P15BHP**, but also in regions where **P15BHP** is highly nonlinear. However, the emulator s_{BL} exhibits small overshoots near discontinuities of **P15BHP**, which explains the somewhat inferior prediction error η_∞ of s_{BL}. The same comment applies to our first test case, see Table 1.

4.3 Numerical Results from High-Dimensional Input

Let us finally present numerical results obtained from high-dimensional input configurations. To this end, we have analyzed responses from output concerning the oil production rate at production well **PRO15** after 13 years, response **P15OPR**, as a function of all seven input variables which were listed at the outset of this section, **AQU**, **GOS**, **WOS**, **VPML**, **VPMH**, **HPML**, and **HPMH**.

We have performed an initial fractional factorial design (FFD) of 79 simulations. To reduce computational complexity, a sequential design is performed in the restricted input space of the three dominating active variables, **HPMH**, **HPML**, and **WOS**. These three main active variables were determined by a sensitivity analysis (using a Pareto plot [9]), on the basis of the 79 initial simulator runs. Further 30 design points were added by sequential design, yielding $m = 109$ design points in total.

Given the high dimension of this test case, $n = 7$, in combination with the small number of design points, $m = 109$, Bayes linear estimation performs remarkably well in terms of prediction quality obtained from its emulator s_{BL}. Indeed, we found $\eta_1 = 4.3$, $\eta_2 = 5.0$, and $\eta_\infty = 13.1$.

5 Conclusion

We have shown the utility of Bayes linear methodology, in combination with sequential adaptive design, for uncertainty evaluation in reservoir forecasting. The resulting Bayes linear estimation has been applied to the PUNQS

test case, a rather simple but fairly realistic and frequently used model problem from reservoir engineering. The performance of the resulting emulator has been compared with that obtained from the response surface methodology (RSM), the basic method of commercial reservoir software, such as COUGAR [2]. We found that the Bayes linear methodology is superior to RSM, especially for highly nonlinear responses. For high-dimensional input data a significant number of more simulator runs need to be included in the initial sequential design. This is illustrated in our previous paper [1].

Acknowledgement

References

1. D. Busby, C.L. Farmer, and A. Iske (2005) Multilevel adaptive design and analysis of computer experiments. Manuscript.
2. The COUGAR project, `http://consortium.ifp.fr/cougar/`.
3. P. S. Craig, M. Goldstein, J. C. Rougier, and A. H. Seheult (2001) Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association* **96**, 717–729.
4. P. S. Craig, M. Goldstein, J. A. Smith, and A. H. Seheult (1998) Constructing partial prior specifications for models of complex physical systems. *The Statistician* **47**, 37–53.
5. C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker (1988) A Bayesian approach to the design and analysis of computer experiments. ORNL Technical Report 6498, National Technical Information Service, Springfield, VA 22161.
6. J. P. Dejean and G. Blanc (1999) Managing uncertainties on production predictions using integrated statistical methods. *SPE Journal*, SPE 56696.
7. M. Goldstein (1998) Bayes linear analysis. *Encyclopaedia of Statistical Sciences*.
8. M. C. Kennedy and A. O'Hagan (2000) Bayesian calibration of computer models. *Journal of the Royal Statistical Society B.* **63**, 425–464.
9. R. H. Myers and D. C. Montgomery (2002) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley.
10. J. Oakley, A. O'Hagan (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society B* **63**, 425–464.
11. The European project PUNQ (**P**roduction Forecasting with **UN**certainty **Q**uantification), `http://www.nitg.tno.nl/punq/`.
12. R Development Core Team (2004) R Foundation for Statistical Computing. `http://www.R-project.org`.
13. J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989) Design and analysis of computer experiments. *Statistical Science* **4**(4), 409–435.
14. A. Saltelli, K. Chan, and M. Scott (2000) *Sensitivity Analysis*, Wiley, New York.