Struct Multidisc Optim (2009) 39:459–475
DOI 10.1007/s00158-008-0346-0

RESEARCH PAPER

Assessing the value of another cycle in Gaussian process surrogate-based optimization

Nestor V. Queipo · Alexander Verde · Salvador Pintos · Raphael T. Haftka

Received: 6 June 2008 / Revised: 20 October 2008 / Accepted: 17 November 2008 / Published online: 8 January 2009

Abstract Surrogate-based optimization (SBO) for engineering design, popular in the optimization of complex engineering systems (e.g., aerospace, automotive, oil industries), proceeds in design cycles. Each cycle consists of the analysis of a number designs, the fitting of a surrogate, optimization based on the surrogate, and exact analysis at the design obtained by the optimization. However, due to time and cost constraints, the design optimization is usually limited to a small number of cycles each with substantial number of simulations (short cycle SBO) and rarely allowed to proceed to convergence. This paper takes a first step towards establishing a statistically rigorous procedure for assessing the merit of investing in another cycle of analysis versus accepting the present best solution. The proposed approach assumes that the set of locations for the next cycle is given, and it relies on: (1) a covariance model obtained from available input/output data, (2) a Gaussian process-based surrogate model, and (3) the *fact* that the predictions in the next cycle are a realization of a Gaussian process with a covariance matrix and mean specified using (1) and (2). Its effectiveness was established using descriptive and inference statistical elements in the context of a well-known test function and the optimization of an alkali-surfactant-polymer flooding of petroleum reservoirs.

Keywords Surrogate-based optimization · Gaussian process · Short cycle optimization

N. V. Queipo (✉) · A. Verde · S. Pintos
Applied Computing Institute, University of Zulia,
Maracaibo, ZU, 4011, Venezuela
e-mail: nqueipo@ica.luz.ve

A. Verde
e-mail: averde@ica.luz.ve

S. Pintos
e-mail: spintos@ica.luz.ve

R. T. Haftka
Aerospace and Mechanical Engineering,
University of Florida, Gainesville, FL 32611, USA
e-mail: haftka@ufl.edu

1 Introduction

Surrogate based optimization (SBO) of computationally demanding simulation-based models has become very popular over the last decade (see Simpson et al. 2002; Li and Padula 2004; Queipo et al. 2005; Wang and Shan 2007). A typical SBO constructs the surrogate based on a number of simulations, estimates the optimum design based on the surrogate, and then performs an exact simulation at that estimated position (*checking phase*). This constitutes one *cycle*. The process is then repeated until resources run out or convergence is established. There has been much progress recently in developing SBO methods with proven convergence (see Rodriguez et al. 1998; Alexandrov 1998), and in the SBO under uncertainty for robust design and reliability-based design optimization as evidenced in the DAKOTA and i-SIGHT optimization frameworks (see Wojtkiewicz et al. 2001; Eldred et al. 2002; Padula et al. 1999; Koch and Gu 2001). However, in many applications, the availability of parallel computation allows us to generate multiple simulations and time limitations dictate a small number of cycles (see Kageyama et al. 2001; Zerpa et al. 2005).

The current frontier of surrogate-based engineering design lacks statistically rigorous procedures for assessing the merit of investing in another cycle of analysis versus accepting the present best solution (PBS). Previous works differ in the infill measure, such as, probability of improvement (PI), and expected improvement (EI), whether the additional cycle includes a single or multiple points, and on the stopping criteria.

Jones et al. (1998) pioneered the use of EI and stopped the search when the maximum EI was less than 1% of the present best solution. Sasena et al. (2002) compared alternative infill sampling plans using a generalized EI measure while stopping the cycles after a fixed number of objective function evaluations. Sobester et al. (2005) used a weighted EI criterion and also limited the cycles to a fixed number of objective function evaluations. Huang et al. (2006) presented a so called augmented EI to address stochastic black box systems and used as stopping criterion a tolerance for the ratio between the maximal EI and the active span of the responses.

Alternatively, Apley et al. (2006) in the context of robust design gave guidelines for additional cycles depending on whether or not the analytical prediction intervals for potential designs overlapped. Forrester and Jones (2008) proposed an EI measure with no user defined parameters and stopped the cycles after a particular target is reached. These works consider the deployment of a single point in each additional cycle. In contrast, clustered approaches for the deployment of multiple points in additional cycles were conducted for PI (Jones 2001) and generalized EI (Ponweiser et al. 2008); the former did not specify an stopping criterion while the latter used a fixed number of objective function evaluations. Using a fixed number of cycles Ginsbourger et al. (2007) gave results for both EI and PI as infill sampling criteria also allowing for multiple points in each additional cycle; two heuristics were used for the EI calculations.

The PI and EI infill measures, as discussed in Jones (2001), have their strengths and limitations; in particular, the PI when coupled with optimization algorithms (searching for points maximizing probability of improvement), the iterates can be shown to be dense (under certain mild assumptions), and can naturally balance local and global searches, but its performance can be sensitive to the target specification. This latter issue can be overcome by identifying PI estimates for alternative targets or by using the target-free EI.

Note that no significant effort has been made to assess the accuracy and statistical significance of the infill measure estimates. This work alleviates this shortcoming providing rigorous statistical estimates of PI (given the locations in the next cycle), and discusses the accuracy and statistical significance of the cited estimates using both analytical and industrial case studies.

The Gaussian processes (GP) perspective to surrogate modeling has a long history in the field of statistics and will prove to be useful in this context. Just as a Gaussian distribution is specified by its mean and a covariance matrix, a Gaussian process is specified by a mean and a covariance model; here, the mean is a function of the location in the model input space, and the covariance is a function expressing how correlated the model output values are at two locations. GP are frequently used for problems of regression (e.g., kriging) and classification and are closely related to a variety of surrogate modeling approaches including neural networks (see Neal 1996), kriging (see Cressie 1993; Chiles and Delfiner 1999), generalized radial basis functions (see Poggio and Girosi 1989), and kernel methods (see Lowe 1995). Rasmussen (see Rasmussen 1996) conducted a comparison of GP regression with several other state of the art methods on a number of problems and, in general, found its performance comparable or superior to most methods. A comparison of GP modeling versus the response surface method is available in Hollingsworth and Mavris (2003).

This paper presents a methodology to address an important step in the decision whether to undertake another cycle. This step is calculating the probability of improving the present best solution beyond a target at a given set of points. The issues of how to select this set of points, and how to account for additional improvement at the optimum based on the updated surrogate are not addressed. The methodology for this step relies on three components: (1) a covariance model (structure and parameters) obtained from available input/output data, (2) a Gaussian process-based surrogate model, and (3) the fact that the predictions in the next cycle are a realization of a Gaussian distribution with a covariance matrix and mean specified using (1) and (2). The methodology is validated using a well-known analytical test function (i.e., F1), and evaluated in the surrogate-based modeling of a field scale alkali-surfactant-polymer (ASP) enhanced oil recovery (EOR) process. ASP flooding is the most promising EOR solution for one of the greatest challenges facing

the oil industry worldwide: after conventional water flooding the residual oil (drops trapped by capillary forces) in reservoirs around the world is likely to be around 70% of the original oil in place (see Dosher and Wise 1976; Lake 1989).

The remainder of the paper is structured as follows: problem statement (Section 2), solution approach (Section 3), evaluation strategy (Section 4), case studies (Section 5), results and discussion (Section 6), and summary and conclusions (Section 7).

2 Problem statement

In the context of surrogate-based optimization, given a surrogate model (built from a set of training points), an exact simulation at the surrogate-based optimum (a cycle); if another cycle is undertaken, *what is the probability of improving the present best solution (PBS) beyond a given target at one or more of arbitrary given prediction sites?* This problem is a key building block toward answering the more general question of whether or not an additional cycle should be undertaken in surrogate-based optimization. This more general question would require finding the set of locations that maximize the probability of improvement and deciding whether or not to conduct another cycle based on the outcome of the optimization procedure and the simulation at the point predicted by the surrogate. As an illustration of the problem of interest, Fig. 1 shows a kriging-based model of a test function (i.e., denoted as F1), a set of prediction sites, the present best solution, and a target-T. Given a probability estimate we also need to establish the accuracy and statistical significance of the probability of improvement estimates.

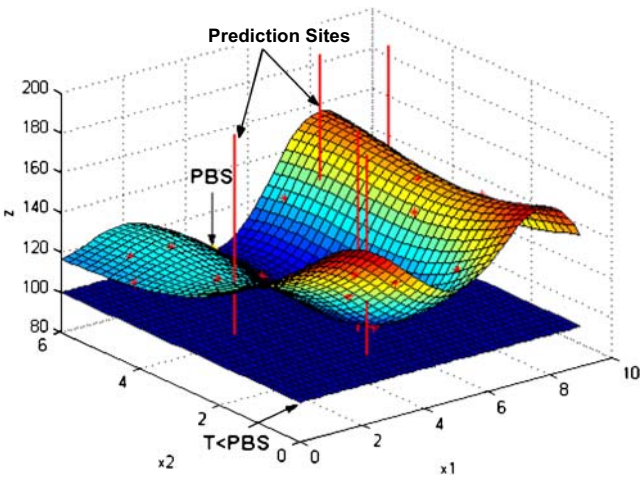

Fig. 1 A Kriging-based model of the F1 test function with a set of five prediction sites where the probability of improving a specified target is sought

Under a GP perspective (see Cressie 1993; Sacks et al. 1989; Williams 1998), the problem of interest can be mathematically formulated as follows. Considering $Z = (Z_t, Z_p)^T$ and $\mathrm{Cov}(Z) = (\Sigma_{tt,} \Sigma_{tp}; \Sigma_{pt}, \Sigma_{pp})$, the points in the next cycle in surrogate-based optimization can be seen as a realization of the following Gaussian distribution (see Rao 2002) with μ being the process mean value:

$$N_{Z_p|Z_t}\left\{\mu + \sum_{pt}\sum_{tt}^{-1}(Z_t - \mu),\ \sum_{pp} - \sum_{pt}\sum_{tt}^{-1}\sum_{tp}\right\} \quad (1)$$

where: Z_t denotes training points plus the exact simulation at the surrogate-based optimum (checking phase), Z_p represents prediction data (output) sets, Σ specify the covariances of the components in vectors Z_p and Z_t, and $N_{Z_p|Z_t}$ is a multivariate normal distribution representing the conditional probability distribution of Z_p given Z_t. Note that the terms in brackets represent the conditional mean at the prediction sites and the conditional covariance matrix of Z_p given Z_t, respectively. The components in the variance and covariance matrices (denoted by Σ) can be calculated by identifying a covariance function using the input/output training data (X_t, Z_t); the general form of the covariance function expresses the idea that nearby inputs will have highly correlated outputs. The GP perspective can be extended to account for scenarios where the conditional mean at the prediction sites is estimated using a variety of surrogate modeling approaches while preserving the conditional covariance matrix of Z_p given Z_t (1).

In this context, the problem of interest is then to calculate the probability that a target T can be met or surpassed by at least one of the components of Z_p given a set of training points in Z_t and exact simulation (checking phase) and to assess the accuracy and statistical significance of the probability of improvement estimates.

3 Solution approach

Given the previously cited GP perspective, the solution approach relies on three components: (1) a covari-

ance model (structure and parameters) obtained from available input/output data (this issue is discussed at the end of this section), (2) a Gaussian process-based surrogate model such as those provided by kriging, and (3) an arbitrary set of P prediction sites and a target (T) corresponding to the next cycle in the surrogate-based optimization. Having specified the Gaussian process (i.e., through its mean and covariance matrix), the probability of interest, that is, the probability of having at least one of the prediction sites (Z_{pj}) below the given target, can be calculated as:

$$\text{Prob}\left(\text{at least } Z_{pj} < T \,|\, Z_t\right) = 1 - \text{Prob}\left(Z_p > W \,|\, Z_t\right) \tag{2}$$

where W is a vector with dimension equal to P and with all its values equal to the target T, that is, $W = \left[\, T \; T \; . \; . \; T \,\right]^t$.

Furthermore, a GP is a stochastic process for which any finite set of outputs (e.g., predictions) has a joint multivariate Gaussian distribution. Hence, considering the symmetry of the multivariate Gaussian density function with respect to its mean value μ (Fig. 2), we can write:

$$\text{Prob}\left(Z_p > W \,|\, Z_t\right) = \text{Prob}\left(Z_p \leq 2\mu - W \,|\, Z_t\right) \tag{3}$$

This transformation is required because we can compute the right hand side of the previous equation using well known algorithms for evaluating Gaussian cumulative probability distributions in high dimensions (See Genz 1993).

More precisely, the solution approach includes the following steps:

1. Construct vector Z_t from the training data. It includes the output values in the data used to construct the surrogate model
2. Identify a covariance model for the Gaussian process using the training data (X_t, Z_t). This issue will be fully discussed later in this section
3. Using the covariance model identified in Step 2, calculate the covariance matrices denoted as: Σ_{tt}, Σ_{tp}, Σ_{pt}, Σ_{pp}
4. Compute the conditional covariance matrix of Z_p given Z_t, that is:

$$\sum_{pp} \Big| Z_t = \sum_{pp} - \sum_{pt} \cdot \sum_{tt}^{-1} \cdot \sum_{tp} \tag{4}$$

5. Create a mean vector $\hat{\mu}$ equal to the surrogate model predictions at the P prediction sites; in the case of simple kriging (assuming a trend equal to zero) the mean vector can be expressed as:

$$\hat{\mu} = \mu + \sum_{pt} \cdot \sum_{tt}^{-1} \cdot (Z_t - \mu) \tag{5}$$

In other *forms of kriging*, the mean at any prediction site would be *the trend value at that location*.
6. Establish a desired target T whose probability of improving in the next cycle is sought and construct a vector $W = \left[\, T \; T \; . \; . \; T \,\right]^t$ with dimension equal to P and components with values equal to T
7. Compute the symmetric vector s with respect to W as:

$$s = 2\hat{\mu} - W \tag{6}$$

8. Compute the value of the multivariate normal cumulative distribution function-CDF corresponding to vector s, namely (Fig. 2): $\text{Prob}\left(Z_p \leq s | Z_t\right)$
9. Calculate the probability of interest as ((2) and (3)):

$$\text{Prob}\left(\text{at least } Z_{pj} < T | Z_t\right) = 1 - \text{Prob}\left(Z_p \leq s | Z_t\right) \tag{7}$$

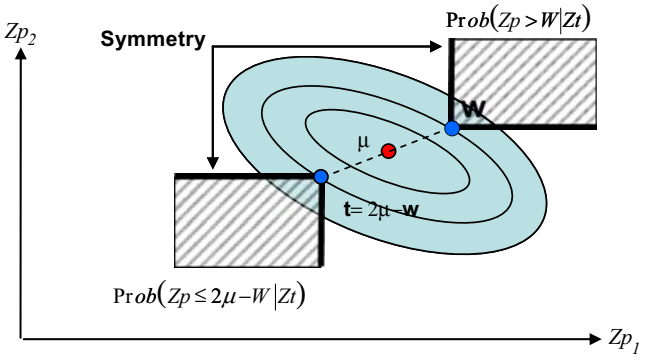

Fig. 2 An illustration of how to estimate Prob $(Z_p > W|Z_t)$ in (2) using multivariate normal cumulative distributions

3.1 Covariance model identification

Identification in this context means to establish the *structure* and *parameters* of the covariance function. As frequently done in the context of surrogate modeling, we assume the GP to be *stationary* in which case the covariance function Cov(z, z') is only a function of the vector h $= x - x'$, namely, Cov(h). The use of such covariance functions is appealing since it makes the prediction invariant under shifts of the origin in the input space, and greatly simplifies the covariance model identification.

From a modeling point of view, we wish to specify a covariance function with a *structure* that embodies our assumptions about the problem (for example that is smooth and continuous). The cited function expresses

the idea that nearby inputs will have highly correlated outputs with some *parameters* (θ) allowing a different distance measure for each input dimension. In addition, we are required to specify a function that will generate positive definite covariance matrices for any set of input points.

One commonly used covariance function for inputs in R^n is:

$$\operatorname{cov}\left(z, z'\right) = \sigma_z^2 \cdot R\left(\theta, x, x'\right) \tag{8}$$

where σ_z^2 is a scale factor and $R\left(\theta, x, x'\right)$ is a correlation function:

$$R\left(\theta, x, x'\right) = \prod_{j=1}^{n} R_j\left(\theta, x_j - x'_j\right) \tag{9}$$

This is simply the product of n correlation functions with a set of parameters θ. Table 1 shows commonly used correlation functions; note that the correlation function does not have to be "Gaussian" and that the parameters model different smoothness behavior in each dimension.

Once the covariance function structure has been set, the *parameters* can be estimated using the training data. There are several approaches for achieving this purpose: (1) maximum likelihood estimates-MLE (see Williams 2002), (2) cross validation (CV), and general cross validation (GCV) methods, as discussed in Wahba (1990), and (3) through variogram modeling (see Cressie 1993). In particular, the MLE approach consists of finding the set of covariance parameters that maximize the log likelihood of the training vector Z_t; the maximum of the likelihood can be estimated using standard optimization routines. The evaluation of the likelihood and its partial derivatives takes $O(n^3)$ operations unless a special structure in the problem can be exploited and can be a difficult problem in high dimensions; approximate methods such as that proposed by Vecchia (1998) have been shown to be useful in such scenarios. A more robust approach for covariance function identification can be made through the so called variogram modeling process from geostatistics.

Table 1 Commonly used correlation functions

Name	$R\left(\theta, x_j - x'_j\right)$
Exponential	$\exp\left(-\theta_j \cdot \left\lvert x_j - x'_j\right\rvert\right)$
Gaussian	$\exp\left(-\theta_j \cdot \left\lvert x_j - x'_j\right\rvert^2\right)$
Exponential – Gaussian	$\exp\left(-\theta_j \cdot \left\lvert x_j - x'_j\right\rvert^{\theta_{n+1}}\right), \quad 0 < \theta_{n+1} \leq 2$

This approach has been limited to low dimensional problems, and extensions to high dimensional problems are not obvious. In any event, there is empirical evidence that even somewhat crude MLEs can lead to useful predictions and quantifications of uncertainty (see Sacks et al. 1989).

4 Evaluation strategy

Multiple (N) experiments are conducted to establish the statistical significance of the results of the proposed approach. For each experiment an additional set of P simulations is conducted and the probability of improving a given target (at the set of P additional locations) is estimated. A successful experiment means that the results of the simulation is below the target T in at least one of the P prediction sites. Then, the goodness of the proposed solution approach is evaluated by:

1. Comparing probabilities of improvement against observed events. Specifically:
   (a) Relative frequency of successful *experiments* vs. average probabilities of improvement
   (b) Average probabilities of improvement associated with successful and unsuccessful experiments.
   (c) Set of instances with probabilities of improvement above average vs. set of instances associated with successful experiments.
2. Through a statistical test, check whether the probabilities of improvement associated with the N experiments are consistent with the observed outcomes.

Note that in the statistical test specified in 2, the N experiments (successful/unsuccessful) can be modeled as the realization of a *Generalized Binomial Distribution* (GBD) and be the subject of a *statistical hypothesis* test with $\alpha = 0.05$ significance level. The random variable for the hypothesis test corresponds to the number of successful events. The adopted model is justified since: (1) the experiments can be considered either successful (1) or unsuccessful (0), (2) the experiments are independent, and (2) each experiment has a different probability of success attached to it (in contrast to a constant probability that leads to the standard Binomial distribution).

The statistical test is conducted assuming as a null hypothesis H_0 that the parameters of the GBD are

the probabilities of improvement associated with the N experiments, and through the following steps:

1. Establish the value k of the random variable X counting the number of successful experiments.
2. Given the probabilities of improvement ($p_{1..N}$) establish the corresponding GBD. As an example, Fig. 3a illustrates the probability of improvement for each experiment in a set of 50 experiments, and the Fig. 3b the GBD associated with those probabilities. The probabilities for individual experiments vary from 0.1 to 0.4, and the most probable number of successful experiments is 12. The GBD distribution then assesses the probability of observing k successful experiments in a set of N experiments given the probability of success (improvement) for each experiment
3. Determine the p-value corresponding to the observed value k. The p-value is the probability of observing a number of successful *experiments* as extreme as k given that the null hypothesis H_0 previously cited is true; H_0 states that the parameters of the GBD are the probabilities of improvement associated with the N experiments. For example, considering the Generalized Binomial probability distribution depicted in Fig. 3b, for $k = 9$ the p-value is 0.34 and is calculated as twice (two-sided test) the sum of the probabilities of improvement associated with number of successful experiments lower or equal to 9. For a value of k on the other side of the mean of the distribution (e.g., 15) the same calculation is made but for a number of successful experiments higher or equal than k
4. If the calculated p-value is greater than the selected significance level ($\alpha = 0.05$) then the null hypothesis cannot be rejected, and the observed experiments are declared consistent with the probabilities of improvement calculated using the proposed approach. For the example in Fig. 3, for the selected significance level of 0.05, the minimum and maximum number of successful experiments that will satisfy this test are 7 (p-value $= 0.09$) and 18 (p-value $= 0.05$), respectively.

Note that the statistical test specified in 2 represents an inference made about the population (of more general value) and goes beyond descriptive statistics.

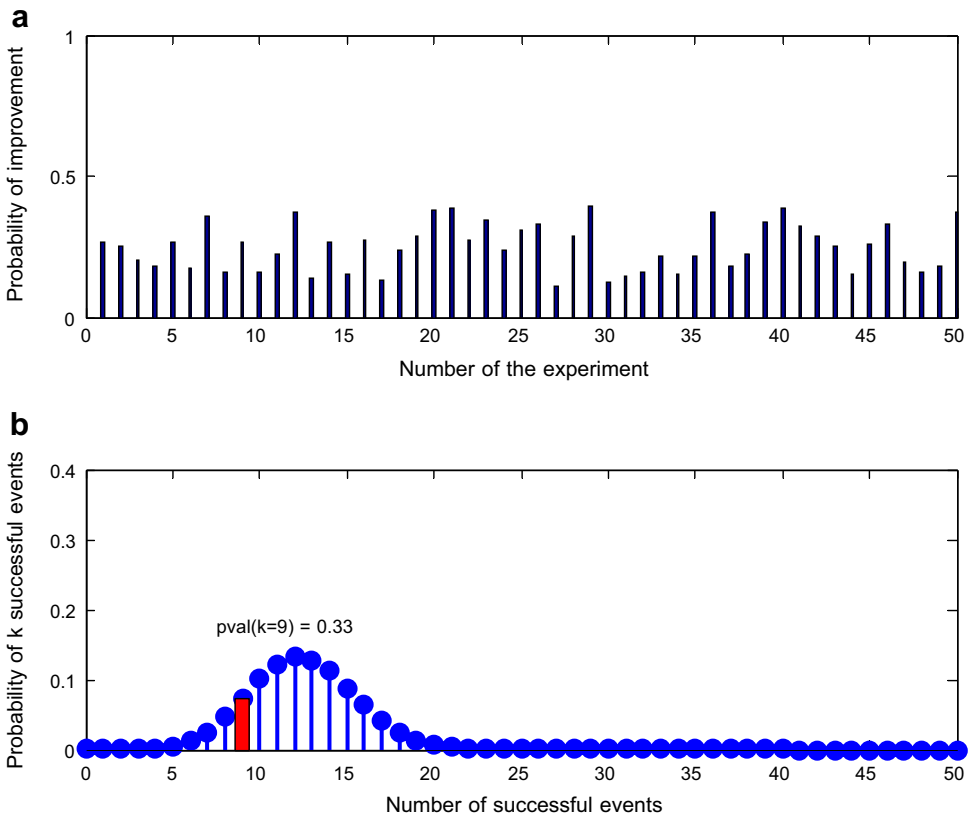

Fig. 3 Illustration of (**a**) probability of improvement for a set of fifty (50) experiments; (**b**) probability of experiencing k successful experiments considering a Generalized Binomial distribution with the probabilities of improvements specified in (**a**) and the p-value associated with the observed result ($k = 9$)

5 Case studies

This section describes the analytical case study (F1 test function) and the industrial case study (Alkali-Surfactant-Polymer enhanced oil recovery optimization—see Zerpa et al. 2005) used to evaluate the proposed solution methodology.

5.1 F1 Test function

This test function (Fig. 4) is expressed as:

$$f(x, y) = (30 + x \sin(x)) \cdot \left(4 + \exp\left(y^2\right)\right) \quad (10)$$

Input space range: $x \in [0, 9]$ and $y \in [0, 6]$

There will be 50 experiments in two different scenarios; that is, $N = 50$ different Latin hypercube designs with 17 $(16 + 1)$ training/checking points, and $P = 5$ prediction sites (scenario I) and 33 $(32 + 1)$ training/checking points, and $P = 5$ prediction sites (scenario II). The target is made equal to the present best solution in each training set, minus 10% of the difference between the present best solution and the global optimum of the function (100.74). Each experiment is characterized by its probability of improvement and an indication of whether or not a successful experiment is observed.

5.2 Alkali-surfactant-polymer (ASP) optimization

In order to increase oil recovery, oil reservoirs may be flooded by a mix of chemicals such as alkali, surfactant and polymer. The problem of interest is to find the values of the design variables, namely, concentration of alkaline, surfactant and polymer, and ASP slug size (expressed in the form of the injection time) that maximize the cumulative oil production over a time horizon. The ranges of the design variables are presented in Table 2. The cumulative oil production is calculated at 487 days using the UTCHEM reservoir simulator (Pope and Nelson 1978; Engelsen et al. 1987), and it is expressed as percentage of the original oil in place (OOIP).

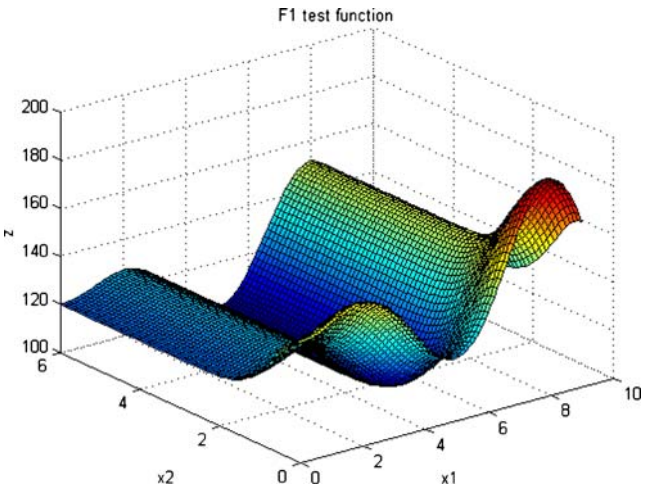

Fig. 4 F1 test function (F1 case study)

The UTCHEM program is a three-dimensional, multiphase, multicomponent reservoir simulator of chemical flooding processes developed at the University of Texas at Austin. The basic governing differential equations consist of: a mass conservation equation for each component, an overall mass conservation equation that determines the pressure (the pressure equation), an energy balance, and Darcy's Law generalized for multiphase flow. The resulting flow equations are solved using a block-centered finite-difference scheme. The solution method is implicit in pressure and explicit in concentration, similar to the well-known IMPES method used in black oil reservoir simulators. A Jacobi conjugate gradient method is used to solve the system of finite difference equations resulted from the discretization of the pressure equation.

As illustrated in Fig. 5, the ASP flooding pilot has an inverted five-spot pattern and a total of 13 vertical wells, among those, nine of those wells are oil producers and four of them are ASP injectors. The reservoir is at a depth of 4,150 ft., has an average initial pressure of 1,770 psi, and its porosity is assumed to be constant throughout the reservoir and equal to 0.3. The OOIP is 395,427 bbl, the crude oil viscosity is 40 cp, the initial brine salinity is 0.0583 meq/ml and the initial brine divalent cation concentration is 0.0025 meq/ml. This is the reference configuration whose details can be found in the sample data archives of the UTCHEM program.

Three flowing phases and 11 components are considered in the numerical simulations. The phases are

Table 2 Design variable restrictions—ASP optimization

Design variable	Range		Units
	Min.	Max.	
Alkaline concentration (Na_2CO_3)	0	0.5898	meq/ml
Surfactant concentration	0.001815	0.01	Vol. fract.
Polymer concentration	0.0487	0.1461	wt.%
Injection time	111	326	Days

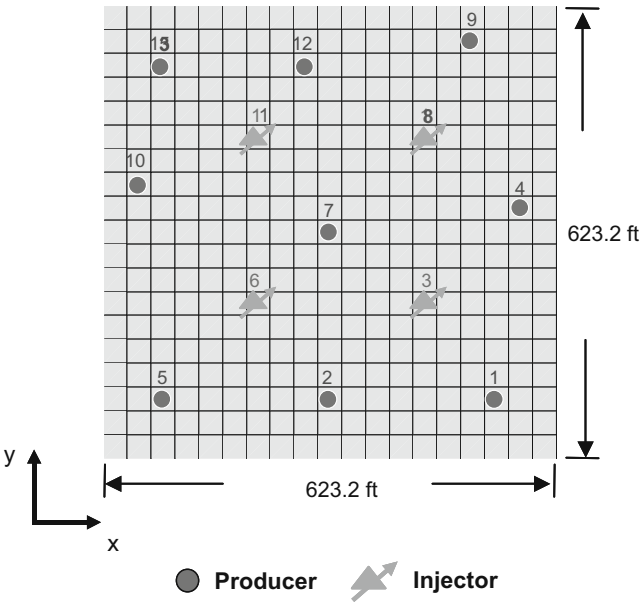

Fig. 5 Well pattern illustration (ASP modeling case study)

water, oil and microemulsion, while the components are water, oil, surfactant, polymer, chloride anions, divalent cations (Ca^{++}, Mg^{++}), carbonate, sodium, hydrogen ion, and oil acid. The ASP interactions are modeled using the reactions: in situ generated surfactant, precipitation and dissolution of minerals, cation exchange with clay and micelle, and chemical adsorption. Note the detailed chemical reaction modeling, and the heterogeneous and multiphase petroleum reservoir under consideration.

Again, there will be 50 experiments; in this case, (50) different Latin hypercube designs with 42 (41 + 1) training/checking points, and $P = 10$ prediction sites. The target is made equal to the best solution in the training set. Note that in both case studies the targets have been modestly set so *that* we improve our chances of observing improvements with a rather limited number of experiments.

In all instances, the selected Gaussian process-based surrogate modeling was ordinary kriging. The modeling was conducted using the Matlab toolbox developed by Lophaven et al. (2002) with a Gaussian correlation function and parameters identified using maximum likelihood principles. The surrogate-based optimization associated with the checking phase was conducted using the DIRECT (Dividing RECTangles) global optimization algorithm, a modified Lipschitzian method, developed by Jones et al. (1993).

Table 4 Calculated average probabilities of improvement for successful and unsuccessful experiments

Objective function	Average probability of successful experiments	Average probability of unsuccessful experiments
F1 (16 + 1)	0.63	0.08
F1 (32 + 1)	0.68	0.03
ASP (41 + 1)	0.70	0.11

6 Results and discussion

Table 3 presents the relative frequency of successful *experiments* (at least one of the prediction sites below target) vs. the calculated average probabilities of improvement for each of 50 experiments in the different case studies, namely: F1 (16 + 1), F1 (32 + 1) and ASP (40 + 1). In the analytical case studies, the agreement is excellent, while in the industrial case study, the observed difference is probably due to the small size of the training data given the number of input variables.

Table 4 shows the calculated probabilities of improvement associated with successful and unsuccessful experiments. The average probabilities of improvement associated with successful *experiments* are higher than 0.6 for the successful experiments and lower than 0.15 for the unsuccessful ones, with even better results (0.63 vs. 0.68; 0.08 vs. 0.03) for a higher sample size in the case study F1(16 + 1) vs. F1(32 + 1).

Table 5 displays the set of instances with calculated probabilities of improvement above and below a 0.5 probability threshold versus the set of successful and unsuccessful experiments for the three case studies under consideration. It is clear that, in all instances, almost all the elements in the set of *unsuccessful* experiments

Table 3 Relative frequency of successful experiments and calculated average probabilities of improvement for the different case studies

Objective function	Successful experiments	Unsuccessful experiments	Relative frequency of successful experiments	Average probability of improvement
F1 (16 + 1)	9	41	0.18	0.18
F1 (32 + 1)	4	46	0.08	0.08
ASP (41 + 1)	3	47	0.06	0.15

Table 5 Set of instances with calculated probabilities of improvement above and below 0.5 vs. set of successful and unsuccessful experiments

	Successful experiments	Unsuccessful experiments	Marginal totals
F1 (16 + 1)			
Experiments with probability above 50%	5	2	7
Experiments with probability below 50%	4	39	43
Marginal totals	9	41	
F1 (32 + 1)			
Experiments with probability above 50%	3	1	4
Experiments with probability below 50%	1	45	46
Marginal totals	4	46	
ASP (41 + 1)			
Experiments with probability above 50%	3	0	3
Experiments with probability below 50%	0	47	47
Marginal totals	3	47	

had probabilities of improvement below the 0.5 probability threshold; namely, 39/41, 45/46, and 47/47 for the F1(16 + 1), F1(32 + 1), and, ASP (40 + 1), respectively. Similarly, for the F1(32 + 1), and, ASP (40 + 1) case studies almost all the elements in the set of *successful* experiments had probabilities of improvement above the 0.5 probability threshold; namely, three fourths and three thirds, respectively.

A statistical measure of the goodness of the results in Table 5 is the balanced error rates (BER), which is the average of the relative errors in predicting successful and unsuccessful events. For example, for the F1 (16+1) case study the BER is equal to $0.5 \times (4/9 + 2/41) = 0.24$. In general, the BER is in the range of approximately 13.5% to 24% for the analytical cases and 0% for the more complex industrial one. Note that better results

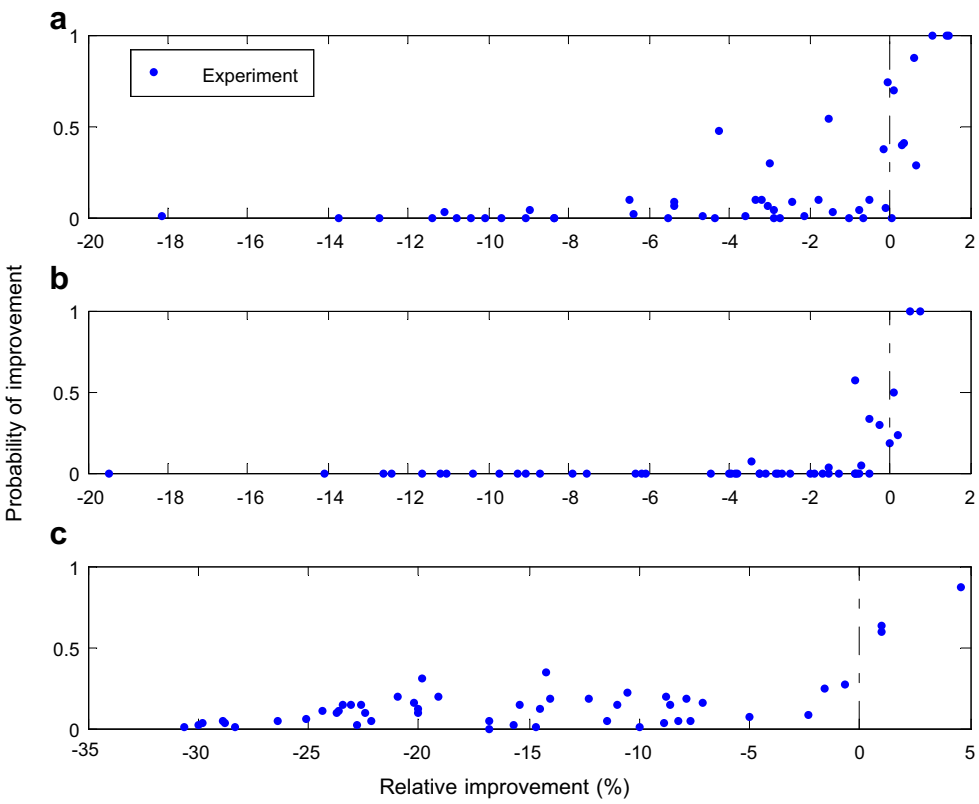

Fig. 6 Probability of improvement vs. actual improvement for each of the 50 experiments in all case studies; **a** F1 case study—(16 + 1), **b** F1 case study—(32 + 1) and ASP case study—(41 + 1)

(13.5% vs. 24%) are obtained for the analytical case study with a higher sample size. In the industrial case study, the BER is equal to 0, even though the average probability of improvement and the relative frequency were on the same order of magnitude but different (0.15 vs. 0.06). The BER = 0 condition only reflects the fact that all successful (unsuccessful) experiments had probability of improvement above (below) the selected threshold and can have associated very different average probability of improvement scenarios. So far in Table 5 we evaluate the prediction performance of the proposed approach using a 0.5 probability threshold. The probability threshold was not set to optimize the results. The optimal probability threshold for deciding whether or not to conduct another cycle is difficult to establish. While a 50% probability threshold seems like a natural choice, the threshold should be selected based on the designer's risk attitude (aversion, neutral, seeking) and the cost in time and resources of carrying another cycle. In addition, the results may be sensitive to alternative covariance functions; however, when using an exponential covariance function in the F1 (16 + 1) case study, the balanced error rate (BER) was also quite good (8%).

Note that this study has focused on assessing the value (probability of improvement) of another cycle. It does not depend on the sampling scheme for the additional cycle in SBO, and no particular effort was made on selecting the prediction sites for optimization. Figure 6 shows the positive relationship between probability of improvement and actual improvement for all case studies with the relative improvements in the analytical and industrial case studies up to 1% and 5%, respectively. While the improvements in the analytical case studies may seem modest, they correspond on average to 48.21% of the maximum theoretically possible (average target 101.30, average optimum 101.01, and optimum value 100.74). In more general scenarios, achieving higher relative improvements is possible if the target and optimum value are significantly apart and an optimization strategy and sampling scheme for the additional cycle is in place.

What follows are the results of the statistical hypothesis tests described in "Section 4" where an inference *is* made about the population and hence goes beyond the descriptive statistics obtained using experiments. Figures 7, 8 and 9 show how in all instances the *p*-values exceed the 5% significance level and hence the

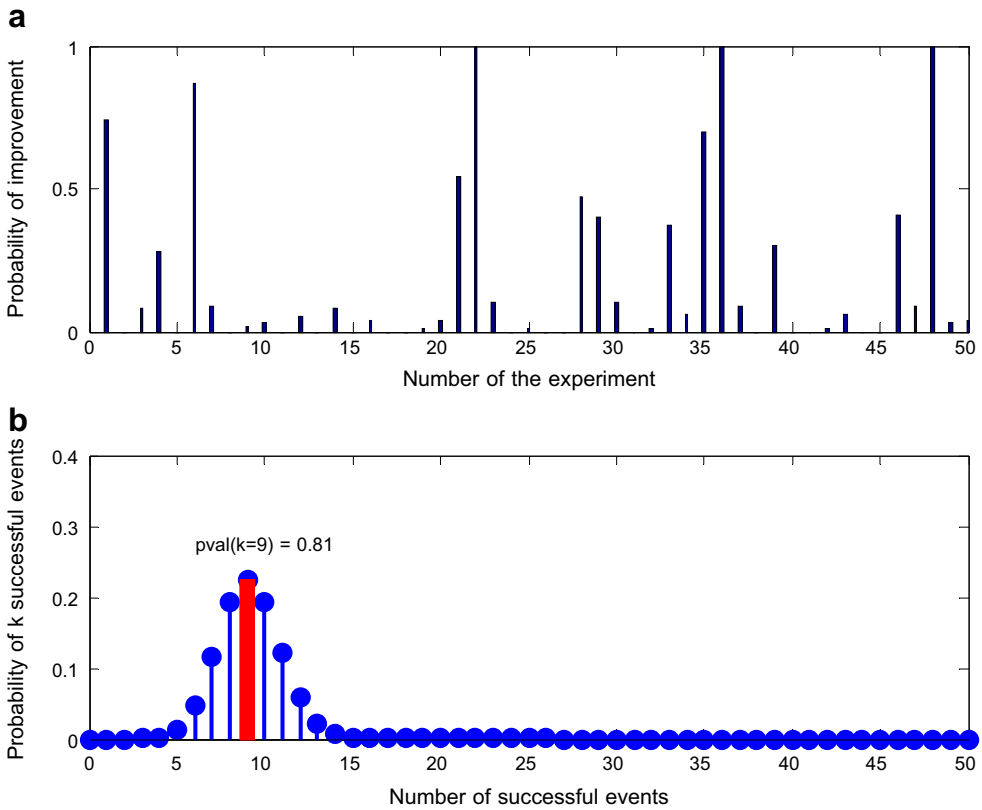

Fig. 7 **a** Calculated probability of improvement for each of the 50 experiments; **b** probability of experiencing k successful experiments and the p-value associated with the observed result ($k = 9$)—F1 case study (16 + 1)

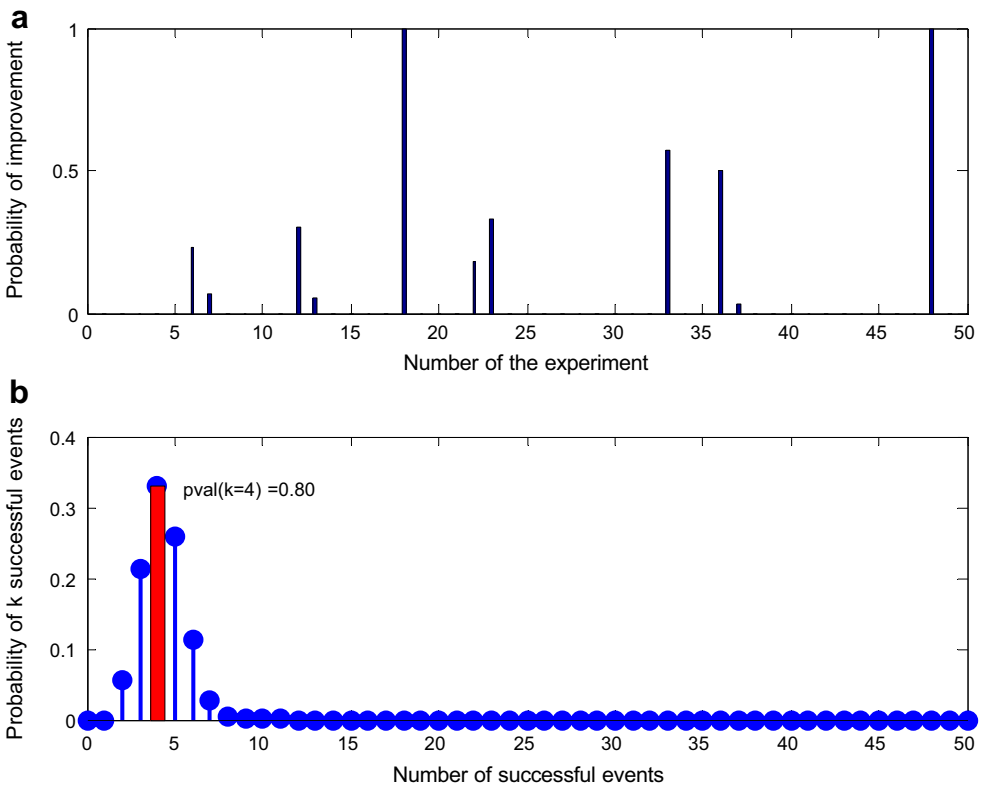

Fig. 8 **a** Calculated probability of improvement for each of the 50 experiments; **b** probability of experiencing k successful events and the p-value associated with the observed result ($k = 4$)—F1 case study ($32 + 1$)

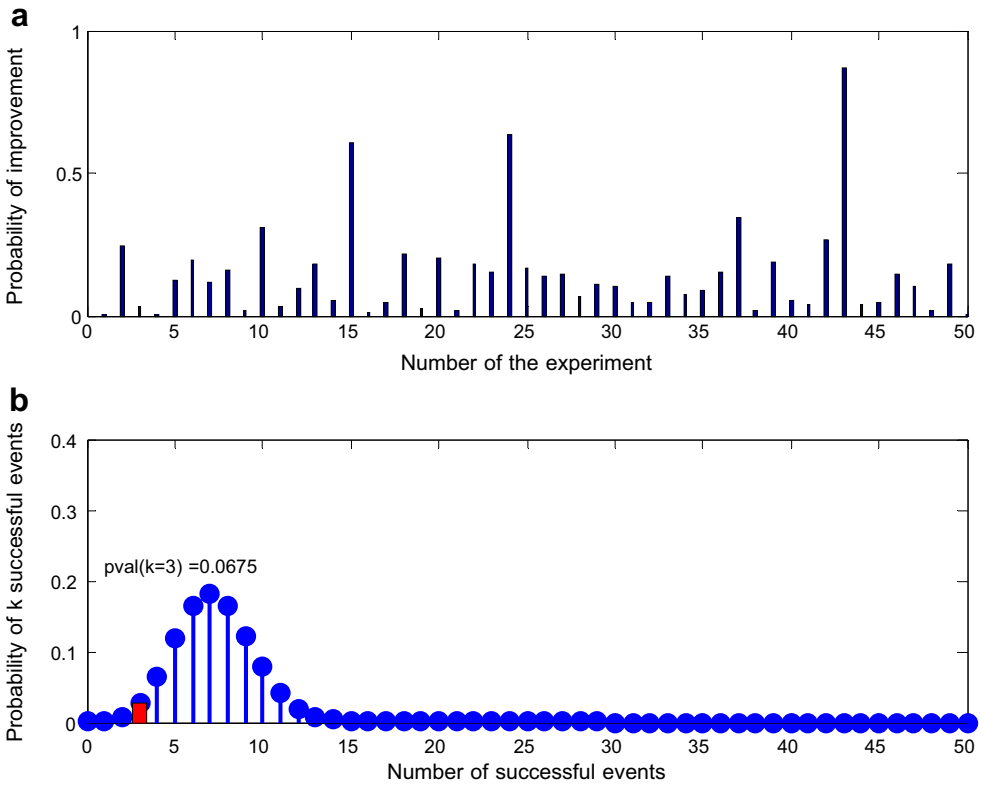

Fig. 9 **a** Calculated probability of improvement for each of the 50 experiments; **b** probability of experiencing successful events and the p-value associated with the observed result ($k = 3$)—ASP case study ($41 + 1$)

calculated probabilities of improvement are considered to be consistent with the experimental results. In particular, the analytical test cases exhibited p-values above 80%.

Detailed results of the experiments for each of the case studies can be found in the Appendix.

7 Summary and conclusions

This paper presented an approach based on the Gaussian process (GP) perspective for evaluating the probability of improvement beyond a target T given a design of experiments for the next cycle. This is a key step in the decision weather to undertake another cycle or accept the present best solution (PBS). The GP perspective to surrogate modeling has a long history in the field of statistics, is frequently used for problems of regression (e.g., kriging) and classification and is closely related to a variety of surrogate modeling approaches including neural networks, kriging, generalized radial basis functions, and kernel methods.

The proposed approach relies on three components: (1) a covariance model (structure and parameters) obtained from available input/output data, (2) a Gaussian process-based surrogate model such as kriging, and (3) the fact that the predictions in the next cycle are a realization of a Gaussian process with a covariance matrix and mean specified using (1) and (2).

The effectiveness of the approach was demonstrated using both descriptive and inference statistics when applied to an analytical case study with two sample sizes and to the surrogate-based optimization of an Alkali-Surfactant-Polymer process, and holds promise to be effective in broader contexts.

Pending issues include extensions to popular alternative surrogate modeling schemes such as polynomial regression, and support vector regression, sound procedures for setting reasonable targets, and the coupling of the proposed approach with formal optimization strategies for selecting the prediction sites in surrogate-based optimization.

Acknowledgements This material is based upon work supported by National Science Foundation under Grant DDM-423280, and the Fondo Nacional de Ciencia, Tecnología e Innovación (FONACIT), Venezuela under Grant F-2005000210. The authors also thank the Center for Petroleum and Geosystems Engineering of The University of Texas at Austin for providing the UTCHEM compositional simulator.

Appendix

Table 6 Results of the different experiments for the F1 case study—(17 training points and five test locations)

No	$\phi 1$	$\phi 2$	μ	σ	p	T	B^a	ymin^b	T- ymin	$(T$- ymin$)/\|T\|$
1	0.39	0.06	130.63	258.81	0.74	102.24	0	102.28	−0.04	0.000
2	0.13	0.04	130.67	641.00	0.00	101.22	0	109.67	−8.45	−0.083
3	1.45	0.04	128.17	265.03	0.08	101.84	0	107.29	−5.45	−0.054
4	0.14	0.07	134.59	260.92	0.28	102.24	1	101.55	0.69	0.007
5	0.18	0.06	136.36	512.80	0.00	101.06	0	114.95	−13.89	−0.137
6	0.28	0.04	134.60	352.39	0.87	101.42	1	100.79	0.63	0.006
7	0.75	0.03	129.34	256.50	0.09	101.39	0	104.62	−3.23	−0.032
8	1.05	0.04	129.43	213.96	0.00	102.24	0	112.54	−10.30	−0.101
9	0.90	0.02	129.43	323.60	0.02	102.24	0	108.75	−6.51	−0.064
10	0.24	0.12	132.25	294.89	0.03	101.84	0	103.27	−1.43	−0.014
11	0.29	0.08	134.12	508.24	0.00	102.14	0	110.68	−8.54	−0.084
12	0.09	0.06	139.57	1,077.89	0.05	101.07	0	101.17	−0.10	−0.001
13	0.42	0.02	130.69	283.98	0.00	100.78	0	101.47	−0.69	−0.007
14	0.40	0.07	132.88	429.47	0.08	100.79	0	103.26	−2.47	−0.025
15	0.26	0.03	127.39	503.87	0.00	100.74	0	113.57	−12.83	−0.127
16	0.83	0.05	131.85	393.36	0.04	102.24	0	111.42	−9.18	−0.090

17	0.09	0.18	139.61	526.41	0.00	101.33	0	112.29	−10.96	−0.108
18	0.29	0.04	128.12	154.30	0.00	100.97	1	100.91	0.06	0.001
19	0.27	0.12	131.60	279.20	0.01	102.04	0	120.58	−18.54	−0.182
20	0.16	0.06	133.55	767.84	0.04	100.78	0	103.69	−2.91	−0.029
21	0.15	0.15	136.77	778.85	0.54	101.01	0	102.56	−1.55	−0.015
22	0.35	0.07	128.79	195.03	1.00	101.87	1	100.80	1.07	0.011
23	0.28	0.04	132.88	247.75	0.10	100.78	0	101.27	−0.49	−0.005
24	0.37	0.04	130.26	205.44	0.00	100.86	0	103.61	−2.75	−0.027
25	0.20	0.26	134.57	333.05	0.01	100.74	0	105.45	−4.71	−0.047
26	0.29	0.18	132.05	261.99	0.00	101.69	0	107.29	−5.60	−0.055
27	0.12	0.05	143.16	1,125.79	0.00	100.92	0	101.96	−1.04	−0.010
28	0.98	0.02	129.27	320.88	0.47	101.21	0	105.53	−4.32	−0.043
29	0.21	0.05	132.34	322.85	0.40	101.60	1	101.27	0.33	0.003
30	0.34	0.04	133.61	382.01	0.10	101.40	0	104.80	−3.40	−0.034
31	0.13	0.09	140.67	580.68	0.00	101.79	0	113.41	−11.62	−0.114
32	0.22	0.17	133.21	407.11	0.01	101.26	0	104.92	−3.66	−0.036
33	0.13	0.08	137.86	858.67	0.37	101.43	0	101.60	−0.17	−0.002
34	0.14	0.07	142.40	789.45	0.06	101.43	0	106.87	−5.44	−0.054
35	0.13	0.15	134.25	451.24	0.70	102.24	1	102.13	0.11	0.001
36	0.18	0.03	137.99	355.36	1.00	102.24	1	100.75	1.49	0.015
37	0.18	0.13	133.59	476.66	0.09	101.03	0	102.84	−1.81	−0.018
38	0.18	0.13	133.89	579.72	0.00	100.80	0	109.94	−9.14	−0.091
39	0.25	0.09	133.25	308.24	0.30	101.00	0	104.03	−3.03	−0.030
40	0.61	0.03	129.64	241.81	0.00	100.76	0	110.52	−9.76	−0.097
41	0.19	0.07	132.18	332.16	0.00	101.09	0	104.00	−2.91	−0.029
42	0.14	0.09	136.33	597.04	0.01	101.42	0	103.57	−2.15	−0.021
43	0.11	0.05	127.68	519.96	0.06	101.64	0	104.75	−3.11	−0.031
44	0.19	0.04	133.61	300.29	0.00	100.98	0	111.51	−10.53	−0.104
45	0.12	0.05	134.61	838.00	0.00	100.87	0	105.28	−4.41	−0.044
46	0.46	0.01	128.07	356.92	0.41	101.41	1	101.03	0.38	0.004
47	0.51	0.04	134.32	441.17	0.09	101.60	0	108.19	−6.59	−0.065
48	0.17	0.05	139.60	632.35	1.00	102.18	1	100.75	1.43	0.014
49	0.29	0.13	131.64	232.34	0.03	101.83	0	113.13	−11.30	−0.111
50	0.34	0.04	132.98	396.54	0.04	102.03	0	102.83	−0.80	−0.008
Average probability					0.18		0.18			

[a]Random variable—number of prediction sites with function values below a given target
[b]Minimum value of test data set

Table 7 Results of the different experiments for the F1 case study—(33 training points and five test locations)

No	$\phi 1$	$\phi 2$	μ	σ	p	T	B	ymin	T– ymin	(T- ymin) / $\lvert T \rvert$
1	0.11	0.13	141.96	946.90	0.00	100.77	0	102.28	−1.51	−0.015
2	0.13	0.12	142.25	1,079.40	0.00	100.89	0	109.67	−8.78	−0.087
3	0.26	0.19	130.03	348.45	0.00	100.89	0	107.29	−6.40	−0.063
4	0.18	0.09	138.10	719.95	0.00	100.74	0	101.55	−0.81	−0.008
5	0.14	0.16	136.34	621.63	0.00	100.75	0	114.95	−14.20	−0.141
6	0.10	0.12	126.50	1,034.09	0.23	101.01	1	100.79	0.22	0.002
7	0.20	0.16	134.64	393.46	0.07	101.11	0	104.62	−3.51	−0.035
8	0.13	0.11	135.97	673.49	0.00	100.78	0	112.54	−11.76	−0.117
9	0.10	0.09	134.17	1,934.99	0.00	100.79	0	108.75	−7.96	−0.079
10	0.20	0.21	128.73	301.56	0.00	100.77	0	103.27	−2.50	−0.025
11	0.12	0.14	136.33	911.48	0.00	100.84	0	110.68	−9.84	−0.098
12	0.13	0.11	138.99	986.25	0.30	100.92	0	101.17	−0.25	−0.002
13	0.14	0.12	134.09	646.80	0.05	100.75	0	101.47	−0.72	−0.007
14	0.14	0.16	134.54	631.26	0.00	101.36	0	103.26	−1.90	−0.019
15	0.10	0.15	134.80	1,028.39	0.00	100.83	0	113.57	−12.74	−0.126
16	0.11	0.11	142.14	977.63	0.00	100.93	0	111.42	−10.49	−0.104
17	0.11	0.18	132.52	729.15	0.00	100.98	0	112.29	−11.31	−0.112
18	0.16	0.14	133.68	422.90	1.00	101.66	1	100.91	0.75	0.007
19	0.10	0.13	135.52	1,021.26	0.00	100.91	0	120.58	−19.67	−0.195
20	0.15	0.09	141.45	773.68	0.00	100.97	0	103.69	−2.72	−0.027
21	0.09	0.10	138.88	1,514.68	0.00	100.87	0	102.56	−1.69	−0.017
22	0.11	0.13	136.92	779.14	0.18	100.78	0	100.80	−0.02	0.000
23	0.17	0.15	133.45	517.99	0.33	100.75	0	101.27	−0.52	−0.005
24	0.16	0.16	132.41	524.35	0.00	100.77	0	103.61	−2.84	−0.028
25	0.18	0.17	132.48	374.00	0.00	100.96	0	105.45	−4.49	−0.044
26	0.09	0.10	134.91	2,194.36	0.00	101.02	0	107.29	−6.27	−0.062
27	0.12	0.09	141.34	740.87	0.00	101.17	0	101.96	−0.79	−0.008
28	0.14	0.15	129.49	499.24	0.00	102.19	0	105.53	−3.34	−0.033
29	0.13	0.11	132.51	534.84	0.00	100.74	0	101.27	−0.53	−0.005
30	0.21	0.07	133.41	981.69	0.00	100.98	0	104.80	−3.82	−0.038
31	0.19	0.10	139.27	660.14	0.00	100.89	0	113.41	−12.52	−0.124
32	0.15	0.13	132.48	880.77	0.00	100.90	0	104.92	−4.02	−0.040
33	0.14	0.10	133.95	598.24	0.57	100.75	0	101.60	−0.85	−0.008
34	0.14	0.15	130.08	732.70	0.00	100.76	0	106.87	−6.11	−0.061
35	0.10	0.11	130.23	1,301.35	0.00	100.86	0	102.13	−1.27	−0.013
36	0.20	0.18	132.42	372.22	0.50	100.84	1	100.75	0.09	0.001
37	0.12	0.13	136.68	788.26	0.03	101.30	0	102.84	−1.54	−0.015
38	0.11	0.11	141.08	866.61	0.00	102.24	0	109.94	−7.70	−0.075
39	0.07	0.09	148.63	3,129.43	0.00	100.77	0	104.03	−3.26	−0.032
40	0.10	0.11	149.11	1,564.48	0.00	101.15	0	110.52	−9.37	−0.093
41	0.11	0.14	137.58	809.37	0.00	100.88	0	104.00	−3.12	−0.031
42	0.12	0.14	135.64	661.75	0.00	100.75	0	103.57	−2.82	−0.028
43	0.13	0.14	131.99	580.19	0.00	100.88	0	104.75	−3.87	−0.038
44	0.19	0.14	135.87	482.97	0.00	102.24	0	111.51	−9.27	−0.091
45	0.14	0.12	138.72	578.04	0.00	101.29	0	105.28	−3.99	−0.039
46	0.16	0.14	100.01	800.23	0.00	101.03	0	101.90	−0.87	−0.009
47	0.13	0.20	140.87	400.13	0.00	108.28	0	109.19	−0.91	−0.008
48	0.12	0.14	137.08	861.71	1.00	101.24	1	100.75	0.49	0.005
49	0.27	0.10	135.12	451.71	0.00	101.85	0	113.13	−11.28	−0.111
50	0.15	0.16	136.37	518.85	0.00	100.82	0	102.83	−2.01	−0.020
Average probability					0.08		0.08			

Table 8 Results of the different experiments for the ASP case study—(41 training points and ten test locations)

No	$\phi 1$	$\phi 2$	$\phi 3$	$\phi 4$	μ	σ	p	T	B	ymin	T- ymin	(T– ymin) / $\lvert T \rvert$
1	0.00	0.15	1,327.85	0.00	−24.85	11.54	0.01	−34.19	0	−30.77	−3.42	0.100
2	0.00	40,659.42	2,616.62	0.00	−23.50	13.51	0.25	−33.65	0	−33.12	−0.52	0.015
3	17.45	0.01	0.19	119.86	−24.32	14.73	0.03	−34.66	0	−24.32	−10.34	0.298
4	0.08	1.91	504.84	0.01	−24.03	13.28	0.01	−34.05	0	−29.06	−5.00	0.147
5	0.18	1.48	1.59	11.78	−23.73	10.47	0.12	−30.99	0	−24.79	−6.19	0.200
6	61.64	0.48	0.06	0.00	−23.91	13.88	0.20	−31.58	0	−28.80	−2.78	0.088
7	61.40	0.00	0.02	0.13	−24.13	15.24	0.12	−33.19	0	−28.39	−4.80	0.145
8	0.00	1.43	543.21	0.27	−24.14	13.02	0.16	−31.25	0	−29.03	−2.22	0.071
9	0.03	2.27	2.62	169.40	−24.05	13.05	0.02	−34.34	0	−24.05	−10.29	0.300
10	48.06	0.01	7.52	3.09	−23.59	11.09	0.31	−29.60	0	−23.72	−5.88	0.199
11	7.22	0.01	1,463.86	0.00	−24.04	10.35	0.03	−31.89	0	−29.05	−2.83	0.089
12	0.00	0.00	60.35	0.78	−24.07	8.76	0.10	−30.62	0	−24.48	−6.15	0.201
13	45.88	0.06	0.01	0.02	−24.36	13.61	0.18	−30.76	0	−26.45	−4.31	0.140
14	1.99	1.78	0.08	37.95	−24.00	14.86	0.05	−33.73	0	−24.00	−9.73	0.288
15	80.06	0.12	318.10	0.00	−23.33	12.91	0.60	−32.29	1	−32.60	0.31	−0.010
16	1.99	1.78	0.08	37.95	−24.01	11.82	0.01	−34.60	0	−24.01	−10.59	0.306
17	0.00	3.95	67,952.16	0.00	−23.83	8.15	0.04	−29.56	0	−27.12	−2.44	0.083
18	0.00	2,772.76	985.98	0.00	−24.44	10.38	0.22	−30.20	0	−27.01	−3.20	0.106
19	1.99	1.78	0.08	37.95	−24.00	12.25	0.03	−33.69	0	−24.00	−9.69	0.288
20	1.38	1.31	0.30	5.61	−23.86	9.91	0.20	−30.23	0	−24.45	−5.78	0.191
21	13.23	3.22	269.37	0.00	−23.16	9.26	0.02	−31.09	0	−26.20	−4.89	0.157
22	0.00	134.13	263.20	0.04	−23.57	10.25	0.18	−30.88	0	−27.10	−3.78	0.122
23	1.23	16.02	2,014.59	0.00	−24.09	8.95	0.15	−30.40	0	−27.79	−2.61	0.086
24	0.00	11.35	1,626.07	0.01	−23.69	10.87	0.63	−30.97	1	−31.27	0.30	−0.010
25	2.62	2.27	0.03	169.40	−24.00	8.36	0.16	−30.08	0	−24.00	−6.08	0.202
26	1.66	1.53	0.16	14.85	−24.27	13.49	0.14	−33.38	0	−29.71	−3.67	0.110
27	17.45	0.01	0.19	119.86	−23.87	11.29	0.15	−31.01	0	−23.87	−7.14	0.230
28	16.58	9,565.86	824.54	0.00	−23.91	6.35	0.07	−30.58	0	−29.05	−1.52	0.050
29	1.58	1.47	0.19	11.48	−24.35	11.45	0.11	−31.90	0	−24.36	−7.54	0.236
30	0.03	2.27	2.62	169.40	−23.83	11.32	0.11	−31.52	0	−23.83	−7.69	0.244
31	0.00	0.04	597.22	0.01	−24.14	11.92	0.04	−32.76	0	−29.00	−3.76	0.115
32	0.00	4,373.66	1,574.19	0.00	−23.96	9.19	0.05	−32.17	0	−29.71	−2.46	0.076
33	682.33	24.27	0.06	0.00	−23.84	14.47	0.14	−32.04	0	−27.11	−4.92	0.154
34	0.02	72.01	2,954.19	0.00	−24.43	12.17	0.08	−31.78	0	−31.05	−0.72	0.023
35	1.99	1.78	0.08	37.95	−24.03	10.07	0.09	−31.51	0	−24.03	−7.48	0.237
36	9.06	0.08	0.03	133.12	−24.17	12.12	0.15	−31.58	0	−24.17	−7.41	0.235
37	1.99	1.78	0.08	37.95	−23.80	7.35	0.34	−28.60	0	−24.54	−4.06	0.142
38	2.62	2.27	0.03	169.40	−23.89	10.10	0.01	−33.33	0	−23.89	−9.45	0.284
39	4.23	3.39	0.01	2,094.73	−23.98	9.62	0.19	−30.33	0	−23.98	−6.35	0.209
40	2.12	1.89	0.06	54.15	−24.10	10.51	0.06	−32.32	0	−24.22	−8.10	0.251
41	0.03	52.42	395.32	0.04	−23.78	10.76	0.04	−31.67	0	−24.66	−7.01	0.221
42	0.00	23.07	22,017.16	0.00	−23.71	12.43	0.27	−31.50	0	−31.28	−0.22	0.007
43	0.00	17,432.39	2,546.31	0.00	−24.17	11.69	0.87	−30.62	1	−32.02	1.40	−0.046
44	1.66	0.17	0.16	14.85	−24.17	12.06	0.04	−33.17	0	−27.59	−5.58	0.168
45	1.99	1.78	0.08	37.95	−24.02	10.78	0.04	−32.64	0	−24.02	−8.61	0.264
46	1,207.98	0.00	0.22	1,396.28	−24.04	10.56	0.14	−31.07	0	−24.04	−7.03	0.226
47	9.06	0.08	0.03	133.12	−23.81	10.69	0.10	−31.28	0	−24.27	−7.01	0.224
48	1.58	1.47	0.19	11.48	−23.94	12.33	0.02	−34.26	0	−26.44	−7.82	0.228
49	1.81	214.68	0.00	0.05	−23.95	11.16	0.18	−30.98	0	−28.55	−2.42	0.078
50	0.70	10.73	814.59	0.00	−24.20	12.29	0.00	−35.46	0	−29.51	−5.95	0.168
Average probability							0.15		0.06			

References

Alexandrov N (1998) A trust region framework for managing the use of approximation models in optimization. Struct Optim 15(1):16–23

Apley DW, Liu J, Chen W (2006) Understanding the effects of model uncertainty in robust design with computer experiments. J Mech Des 128:945, July

Chiles JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley-Interscience, Hoboken

Cressie N (1993) Statistics for spatial data. Wiley, New York

Dosher TM, Wise FA (1976) Enhanced oil recovery potential. An estimate. Paper SPE 5800, J Petroleum Technology, pp 575

Eldred MS, Giunta AA, Wojtkiewicz SF Jr., Trucano TG (2002) Formulations for surrogate-based optimization under uncertainty. In: 9th AIAA/ISSMO Symposium on multidisciplinary analysis and optimization, Atlanta, GA, AIAA-2002–5585

Engelsen S, Lake LW, Lin EC, Ohno T, Pope GA, Camilleri D, Sepehrnoori K (1987) Description of an improved compositional micellar/polymer simulator. Paper SPE 13967, SPE Reservoir Engineering, pp 427–432

Forrester A, Jones D (2008) Global optimization of deceptive functions with spare sampling. In: 12th AIAA/ISSMO multidisciplinary analysis and optimization conference, Victoria, British Columbia, Canada, 10–12 September

Genz A (1993) A comparison of methods for numerical computation of multivariate normal probabilities. Comput Sci Stat 25:400–405

Ginsbourger D, Le Riche R, Carraro L (2007) Multi-points criterion for deterministic parallel global optimization based on Kriging. In: Intl. Conf. on Nonconvex Programming, NCP07, Rouen, France, December

Hollingsworth P, Mavris D (2003) A technique for use of Gaussian processes in advanced meta-modeling. Proceedings of the SAE Aerospace Congress and Exhibition, Montreal, Canada

Huang D, Allen TT, Notz WI, Zeng N (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models. J Glob Optim 34:441–466

Jones D (2001) A taxonomy of global optimization methods based on response surfaces. J Glob Optim 21:345–383

Jones DR, Perttunen CD, Stuckman BE (1993) Lipschitzian optimization without the Lipschitz constant. J Optim Theory Appl 79(1):157–181

Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. J Glob Optim 13:455–492

Kageyama Y, Yu Q, Shiratori M (2001) Robust and optimal parameter design using SDSS and reliability design and application to hydraulic servo valve. In: Hernandez S, Brebbia CA (eds) Computer aided optimum design of structures VII. MIT Press, Southampton, pp 13–22

Koch PN, Gu L (2001) Addressing uncertainty using the iSIGHT probabilistic design environment. First Annual Probabilistic Methods Conference, Newport Beach

Lake L (1989) Enhanced oil recovery. Prentice Hall, Englewood Cliffs, p 408

Li W, Padula S (2004) Approximation methods for conceptual design of complex systems. In: 11th international conference on approximation theory

Lophaven S, Nielsen H, Sondergaard J (2002) DACE: A Matlab kriging toolbox, version 2.0. Informatics and Mathematical Modeling, Technical University of Denmark. http://www.imm.dtu.dk/~hbn/dace/

Lowe D (1995) Similarity metric learning for a variable kernel classifier. Neural Comput 7:72–85

Neal R (1996) Bayesian learning for neural networks. Lecture notes in statistics, vol 118. Springer, New York

Padula SL, Korte JJ, Dunn HJ, Salas A (1999) Multidisciplinary optimization branch experience using iSIGHT software. NASA/TM-1999-209714, Langley Research Center, Hampton

Poggio T, Girosi F (1989) A theory of networks for approximation and learning. Technical report A.I., 1140, M.I.T

Ponweiser W, Wagner T, Vincze M (2008) Clustered multiple generalized expected improvement: a novel infill sampling criterion for surrogate models. In: Proceedings of the IEEE congress on evolutionary computation (CEC 2008), June 1–6, Hong Kong, IEEE, pp 3514–3521. ISBN 978-1-4244-1823-7

Pope GA, Nelson RC (1978) A chemical flooding compositional simulator. Paper SPE 6725, Society of Petroleum Engineers

Queipo N, Haftka R, Shyy W, Goel T, Vaidyanathan R, Tucker PK (2005) Surrogate-based analysis and optimization. J Prog Aerosp Sci 41:1–28

Rao CR (2002) Linear statistical inference and its applications, 2nd edn. Wiley, New York, pp 516–533

Rasmussen C (1996) Evaluation of Gaussian processes and other methods for non-linear regression. PhD Thesis. Department of Computer Science, University of Toronto

Rodriguez JF, Renaud JE, Watson LT (1998) Trust region augmented Lagrangian methods for sequential response surface approximation and optimization. ASME J Mech Design 120:58–66

Sacks J, Welch W, Mitchell T, Wynn H (1989) Design and analysis of computer experiments. Stat Sci 4:409–423

Sasena MJ, Papalambros P, Goovaerts P (2002) Exploration of metamodeling sampling criteria for constrained global optimization. Eng Optim 34(3):263–278, January

Simpson TW, Booker AJ, Ghosh D, Giunta AA, Koch PN, Yang R-J (2002) Approximation methods in multidisciplinary analysis and optimization: a panel discussion. In: 3rd ISSMO/AIAA internet conference on approximations in optimization

Sobester A, Leary SJ, Keane AJ (2005) On the design of optimization strategies based on global response surface approximation models. J Glob Optim 33:31–59

Vecchia A (1998) Estimation and model identification for continuous spatial processes. Statist Soc Ser B 50:297–312

Wahba G (1990) Spline models for observational data. In: SIAM, CBMS-NSF regional conference series in Applied Mathematics

Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. ASME Trans J Mech Des 129:370–380, April 2007

Williams C (1998) Prediction with Gaussian processes: from linear regression to linear prediction and beyond. Learning in graphical models. Kluwer Academic, Dordrecht, pp 599–621

Williams C (2002) Gaussian processes. In: Arbib M (ed) The handbook of brain theory and neural networks, 2nd edn. MIT Press, Cambridge

Wojtkiewicz SF, Eldred MS, Field RV Jr, Urbina A, Red-Horse JR (2001) A toolkit for uncertainty quantification in large computational engineering models. In: Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference. Seattle, WA, AIAA-2001–1455

Zerpa L, Queipo N, Pintos S, Salager J (2005) An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates. J Pet Sci Eng 47:197–208