Struct Multidisc Optim (2008) 36:247–261
DOI 10.1007/s00158-007-0159-6

RESEARCH PAPER

Toward an optimal ensemble of kernel-based approximations with engineering applications

Egar Sanchez · Salvador Pintos · Nestor V. Queipo

Received: 28 December 2006 / Revised: 11 May 2007 / Accepted: 28 May 2007 / Published online: 11 August 2007

Abstract This paper presents a general approach toward the optimal selection and ensemble (weighted average) of kernel-based approximations to address the issue of model selection. That is, depending on the problem under consideration and loss function, a particular modeling scheme may outperform the others, and, in general, it is not known a priori which one should be selected. The surrogates for the ensemble are chosen based on their performance, favoring non-dominated models, while the weights are adaptive and inversely proportional to estimates of the local prediction variance of the individual surrogates. Using both well-known analytical test functions and, in the surrogate-based modeling of a field scale alkali-surfactant-polymer enhanced oil recovery process, the ensemble of surrogates, in general, outperformed the best individual surrogate and provided among the best predictions throughout the domains of interest.

Keywords Kernel-based approximation · Surrogate-based modeling · Optimal ensemble

1 Introduction

The surrogate-based modeling approach is increasingly popular and has been shown to be useful in the analysis and optimization of computationally expensive simulation-based models in, for example, the *aerospace* (Balabanov et al. 1998; Giunta et al. 1997; Li and Padula 2004; Queipo et al. 2005), *automotive* (Craig et al. 2002; Kurtaran et al. 2002), and *oil industries* (Queipo et al. 2002a,b). Surrogate-based modeling makes reference to the idea of constructing an alternative fast model (surrogate) from numerical simulation data and using it for analysis and optimization purposes. However, practitioners still have to deal with the issue of model selection where, depending on the problem under consideration and loss function (i.e., quadratic, Laplace, ϵ-insensitive), a particular modeling scheme (e.g., polynomial regression, linear splines, Gaussian radial basis functions, or Kriging) may outperform the others, and, in general, it is not known a priori which one should be selected (Jin et al. 2001; Simpson et al. 2001). While there are significant efforts to address the above-referenced issue, practitioners are still looking for guidelines on how to optimally perform model selection.

On the other hand, kernel-based methods (Girosi 1998; Müller et al. 2001) provide the flexibility of generating models under alternative loss functions and, in

This work was supported in part by the Fondo Nacional de Ciencia, Tecnología e Innovación (FONACIT), Venezuela under Grant F-2005000210. N. Q. Author also acknowledges that this material is based upon work supported by National Science Foundation under Grant DDM-423280.

E. Sanchez · S. Pintos · N. V. Queipo (✉)
Applied Computing Institute, Faculty of Engineering,
University of Zulia, Maracaibo 4011, Venezuela
e-mail: nqueipo@ica.org.ve, nqueipo@ica.luz.ve

E. Sanchez
e-mail: esanchez@ica.luz.ve

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

Table 1 Kernel functions associated with a variety of modeling schemes

Kernel	Parametrization
Polynomial order d	$k(\mathbf{x},\mathbf{x}') = \left(\langle \mathbf{x},\mathbf{x}'\rangle + c\right)^d \quad d \in N, c \geq 0$
Spline	$k(\mathbf{x},\mathbf{x}') = 1 + \langle \mathbf{x},\mathbf{x}'\rangle + 1/2 \langle \mathbf{x},\mathbf{x}'\rangle \min(\mathbf{x},\mathbf{x}') - 1/6 \min(\mathbf{x},\mathbf{x}')^3$
B-spline order 2n+1	$k(\mathbf{x},\mathbf{x}') = B_{2n+1}\left(\lVert \mathbf{x}-\mathbf{x}'\rVert\right) B_k = \otimes_{i=1}^{k} \mathbf{I}_{[-1/2,1/2]}$
RBF	$k(\mathbf{x},\mathbf{x}') = exp\left(-\lVert \mathbf{x}-\mathbf{x}'\rVert^2 / 2h^2\right) \quad h > 0$
ERBF	$k(\mathbf{x},\mathbf{x}') = exp\left(-\lVert \mathbf{x}-\mathbf{x}'\rVert / 2h^2\right) \quad h > 0$

particular, support vector regression (SVR) developed by Vapnik (1998) at AT&T Labs in the mid 1990s. It is a rapidly developing field of research, already giving state-of-the-art performances in a variety of applications, which provides a powerful alternative to conduct surrogate-based analysis and optimization. For a discussion of SVR applications in engineering, see for example Clarke et al. (2005) and references therein.

The power of SVR resides in several fronts, such as: (1) *robustness and sparseness of the solution*; the goodness of fit is measured not by the usual quadratic loss function (mean square error) but by a different loss function (ϵ-insensitive) similar to those used in robust statistics (i.e., a way of dealing with deviations from idealized assumptions) and a (2) *flexible and mathematically sound approach*; non-linear regression models (e.g., polynomials, Gaussian radial basis functions, splines, etc.) can be constructed as linear ones by mapping the input data into a so-called feature space, namely, a reproducing kernel Hilbert space (Wahba 2000). The linear models (a single framework) are formulated in terms of dot products in a feature space which can be efficiently calculated using special functions (kernels) associated with the non-linear regression models of interest evaluated in the original space (kernel trick). This framework can also be used with quadratic loss functions, which makes it an ideal setting for ensembles of surrogate-based analysis and optimization.

Previous efforts in the area of model selection have focused on either: (1) *select a particular surrogate from a set of candidates* using, for example, Akaike information criterion (AIC; Buckland et al. 1997; Martin and Simpson 2005), Bayesian information criterion (BIC; Hoeting et al. 1999; Kass and Raftery 1995) or cross-validation methods, or novel techniques based on learning theoretic performance bounds such as the structural risk minimization method (Cherkassky and Ma 2003; Cherkassky et al. 1999) or (2) *build an ensemble of the available surrogates* (weighted average) with weights calculated based on global (Bishop 1995; Goel et al. 2007; Perrone 1994; Perrone and Cooper 1993; e.g., AIC, BIC, MSE) or local (Zerpa et al. 2005) performance measures.

The ensemble of surrogates approach accounts for model selection, and there is evidence that it can provide better average predictive ability than using any single model (e.g, Madigan and Raftery 1994), while the variant of computing the weights using local performance measures (prediction variance) consider the fact that surrogates rank differently throughout the input space. Zerpa et al. (2005) used analytical prediction variance (known to underestimate the true values) as local performance measures, but did not provide a strategy to select surrogates to build the ensemble and was limited to a quadratic loss function.

This paper provides a general approach, automatic, with a reasonable computational cost toward the optimal selection and ensemble (weighted average) of kernel-based models under alternative loss functions, with weights based on empirically estimated prediction variances while promoting diversity among the selected

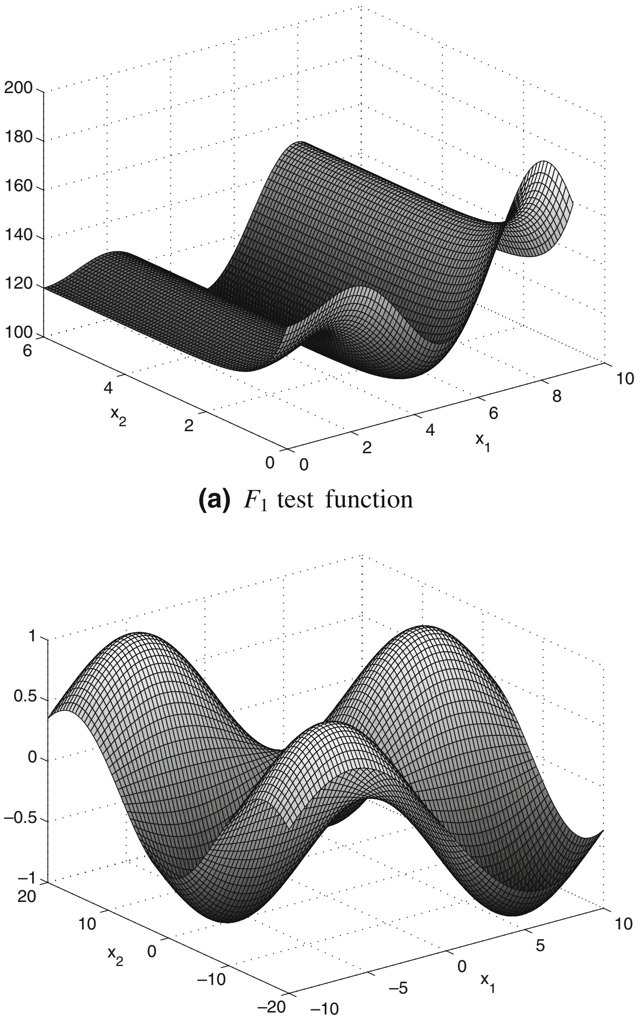
Table 2 Surrogate models under consideration with the ε-insensitive loss function

Kernel	Parameter	Constant C	ε	Total number of models	Models selected
RBF	h_{cm}			15	1
	$0.5h_{\mathrm{cm}}$			15	1
ERBF	h_{cm}	$0.25C_{\mathrm{cm}}$		15	1
	$0.5h_{\mathrm{cm}}$	$0.5C_{\mathrm{cm}}$	0	15	1
Polynimial	Degree 2	$0.75C_{\mathrm{cm}}$	0.05	15	1
	Degree 3	$100C_{\mathrm{cm}}$	0.1	15	1
Spline	Degree 3	$1.50C_{\mathrm{cm}}$		15	1
B-spline	Degree 2			15	1
	Degree 3			15	1

Table 3 Surrogate models under consideration with a Quadratic loss function

Kernel	Parameter	Constant C	Total number of models	Models selected
RBF	h_{cm}		5	1
	$0.5h_{\mathrm{cm}}$		5	1
ERBF	h_{cm}	$0.25C_{\mathrm{cm}}$	5	1
	$0.5h_{\mathrm{cm}}$	$0.50C_{\mathrm{cm}}$	5	1
Polynomial	Degree 2	$0.75C_{\mathrm{cm}}$	5	1
	Degree 3	$1.00C_{\mathrm{cm}}$	5	1
Spline	Degree 3	$1.50C_{\mathrm{cm}}$	5	1
B-spline	Degree 2		5	1
	Degree 3		5	1

models; the latter has been shown to increase the benefits of the ensemble approach (Krogh and Sollich 1997). Its performance is evaluated using both well-known analytical test functions, and, 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 (Dosher and Wise 1976; Lake 1989).



(a) F_1 test function

(b) F_2 test function

Fig. 1 Analytical test functions

2 Problem definition

Given a training sample $E = ((x_h, y_h) : 1 \leq h \leq n)$ of a function $y = f(x)$ defined in $D \subset R^q$, and l kernel-based surrogate models M_i, $1 \leq i \leq l$ constructed from sample E, select a set of m surrogate models and build a weighted average model:

$$Wavg(x) = \sum_{i=1}^{m} \beta_i(x) M_i(x) \tag{1}$$

such that the weighted average model outperforms as many individual surrogates as possible. In the equation above, $\beta_i(x)$ represents the weight of model $M_i(x)$ at location x, and the performance measures are: mean absolute error (ma) $\frac{\sum_{h=1}^{n} abs(y_h - M(x_h))}{n}$, standard deviation (std) $\sqrt{\frac{\sum_{h=1}^{n}(y_h - M(x_h))^2}{n-1}}$, and maximum absolute error (max) $\max abs(y_h - M(x_h))$, with $1 \leq h \leq n$.

3 Solution methodology

It includes the following steps:

1. For each of the case studies, a Latin hypercube sample (sparse) from the model input space is

Table 4 Coefficients used for the evaluation of the analytical test function F_3

a_{ij}							c_i	P_{ij}					
1	10.0	3.0	17.0	3.05	1.7	8.0	1.0	0.1312	0.1696	0.5569	0.0124	0.8283	0.5886
2	0.05	10.0	17.0	0.1	8.0	14.0	1.2	0.2329	0.4135	0.8307	0.3736	0.1004	0.9991
3	3.00	3.5	1.7	10.0	17.0	8.0	3.0	0.2348	0.1451	0.3522	0.2883	0.3047	0.6650
4	17.0	8.0	0.05	10.0	0.1	14.0	3.2	0.4047	0.8828	0.8732	0.5743	0.1091	0.0381

drawn**,** and the corresponding model outputs are calculated.

2. The model input and output values are normalized to the scale $[-1, 1]$.
3. For each of the SVR models differing in kernel and loss function, a set of parameters are specified, namely, C (regularization parameter) and ϵ (size of insensitive zone); the kernels in SVR models are described, for example, as Gaussian with width h or polynomial with degree p (see next section for details). The parameters C and ϵ were identified using cross-validation (k-fold strategy) such that they minimize the mean absolute value of the errors. Specifically, after dividing the data into n/k clusters, each fold is constructed using an element from each of the clusters so it is a representative sample of the model of interest. Note the diversity of the potential members of the ensemble, as all the models differ in either their kernel or loss function. The next section provides an introduction to kernel-based regression.
4. Select the best m models among the set of non-dominated models. The selection criterion is the mean absolute cross-validation error, and non-dominated models make reference to models that provide the lowest error prediction in at least one point in the training data.
5. Using the m models specified in the previous step, a weighted average model is constructed. The adaptive weights $\beta_j(x)$ are inversely proportional to an estimation of the prediction variance $\sigma_j^2(x)$ of M_j at point x. The local prediction variance for each of the models is estimated empirically using the v nearest neighbors of point x. Specifically,

$$\sigma_j^2(x) = \frac{1}{(v-1)} \sum_{h=1}^{v} (y(s_h) - M_j(s_h))^2 \tag{2}$$

where $s_1, s_2, ..., s_v$ are the v nearest neighbors of point x whose corresponding model outputs are $y(s_1)$, $y(s_2)...y(s_v)$. The weight for model M_j is then given as:

$$\beta_j(x) = \frac{\dfrac{1}{\sigma_j^2(x)}}{\sum_{k=1}^{m} \dfrac{1}{\sigma_k^2(x)}} \tag{3}$$

These weights can be shown to be an optimal selection (Bishop 1995) for the case of uncorrelated models in the ensemble. Optimality here makes reference to the best linear unbiased estimator (minimum variance).

4 Kernel-based regression

The kernel-based regression models M_is can be seen as solutions of the following variational problem:

$$\min_{M \in H} Z(M) = \frac{1}{n} \sum_{i=1}^{n} L\left(y_i - M\left(x_i\right)\right) + \lambda \left\| M \right\|_H^2 \tag{4}$$

over some large space of functions H where L and λ denote a particular loss function (e.g., quadratic, Laplace, ϵ-insensitive, and Huber loss functions) and a regularization parameter, respectively. The operator

Table 5 Input variable restrictions (ASP modeling case study)

Input variable	Range		Units
	Min	Max	
Alkaline concentration (Na2CO3)	0	0.5898	meq/ml
Surfactant concentration	0.001815	0.005	Vol. fract.
Polymer concentration	0.0487	0.12	wt%
Injection time	111	326	days

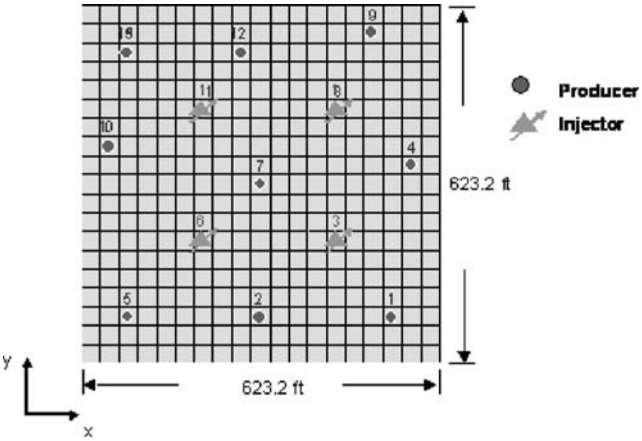


Fig. 2 Reference configuration (ASP modeling case study)

$\|\cdot\|_H^2$ is the Hilbert-space norm which penalizes models that are too complex.

If we restrict ourselves to reproducing kernel Hilbert spaces (RKHS) the variational problem can be formulated as stated in (5).

$$\min_{M \in H} Z(M) = \frac{1}{n} \sum_{i=1}^{n} L\left(y_i - \langle M, K_{x_i} \rangle\right) + \lambda \langle M, M \rangle_H \quad (5)$$

It can be shown that independently of the form of the loss function, the solution of the variational problem can be expressed as:

$$M(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i) \quad (6)$$

where k represents a kernel function. Table 1 shows kernel functions associated with a variety of surrogate modeling schemes.

In particular, if the loss function is quadratic, the coefficients in (6) can be found by solving the following linear system:

$$(n\lambda I + K)\alpha_i = y_i$$

where $\boldsymbol{K}$ denotes the so-called Gram matrix with component K_{ij} denoting $k(x_i, x_j)$, and $\boldsymbol{I}$ representing the identity matrix. Alternatively, if the ϵ-insensitive loss function is used, the coefficients in (6) are found by solving a quadratic programming problem. See Schölkopf and Smola (2002) and Poggio and Smale (2003) for details.

5 Case studies

5.1 General considerations

The solution methodology is evaluated using three well-known (Jin et al. 2001) test functions (F_1, F_2, F_3) with and without noise and a modeling problem in the area of enhanced oil recovery. The test functions with noise consider two noise levels ($\alpha_1 = 0.05$, $\alpha_2 = 0.1$) and a uniform noise distribution U, as specified by the following expression: $F_k\left(1 + \alpha\left(U - \frac{1}{2}\right)\right)$.

Tables 2 and 3 show the models under consideration with quadratic and ϵ-insensitive loss functions and kernels for polynomial, Gaussian radial basis functions, exponential radial basis functions, splines, and B-splines as specified in Table 1. Third- and second-order polynomials and third-degree splines are considered. The ϵ and C values under consideration are 0, 0.05, 0.1 and $0.5\,C_{\mathrm{cm}}$, $0.75\,C_{\mathrm{cm}}$, $1.00\,C_{\mathrm{cm}}$, $1.5\,C_{\mathrm{cm}}$, respectively; the h values are set equal to $1.00h_{\mathrm{cm}}$ and $0.50h_{\mathrm{cm}}$, with C_{cm} and h_{cm} reference values as proposed by Cherkassky and Ma (2004). The kernel-based regression problems are solved using the Matlab support vector machines (SVM) toolbox (Gunn 1998).

The parameter values are selected using cross-validation (k-fold) and 20 training points with $k = 5$ for the analytical F_1 and F_2 test cases (Section 5.2), 60 training points with $k = 10$ for the analytical F_3 test case, 64 training points with $k = 8$ for the ASP

Table 6 Reservoir and fluid properties (ASP modeling case study)

Property	Value	Unit
Reservoir depth	4150 (1265)	ft (m)
OOIP	395,427 (62,868)	bbls(m^3)
Oil viscosity	40	cp
Porosity	0.3	fraction
Average Initial Pressure	1770	psi
Well ratio	0.49 (15)	ft (m)
Skin factor	0.0	adim
Water salinity	C_{Na}^{+2} (0.0583)	meq/ml
	C_{Ca}^{+2} (0.0025)	meq/ml

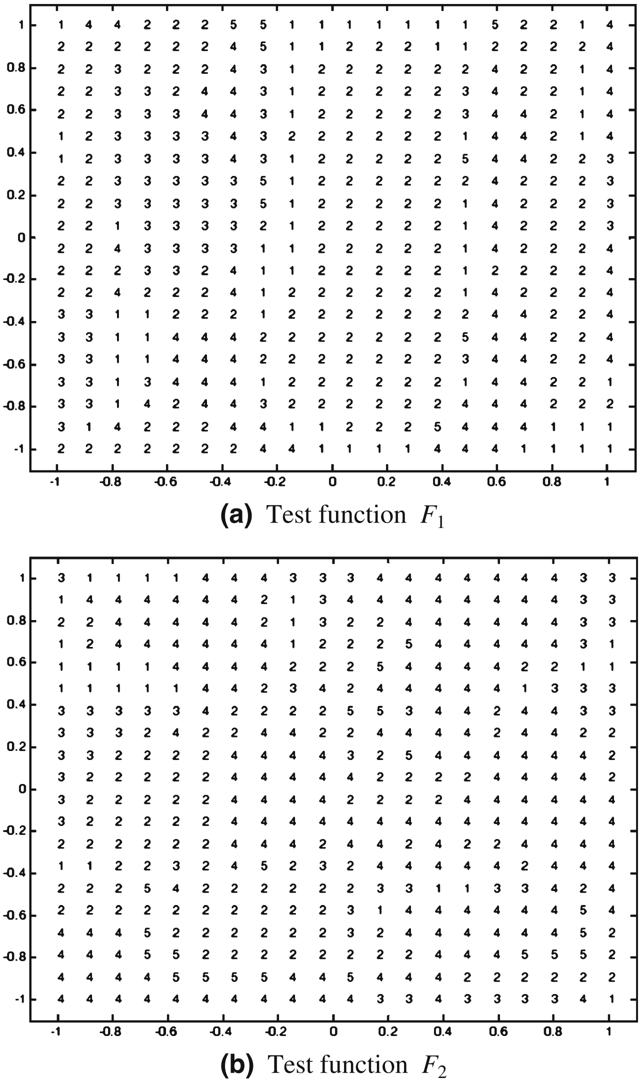
Table 7 Characterization of the models in the ensemble of five models based on loss and kernel functions for the different scenarios. The sequence $X - F_i - Y$ represents the sample, the test function and noise level (if applicable)

Training sample	Loss function		Kernel function				
	ϵ-Insensitive	Quadratic	RBF	ERBF	Poly	Spline	B-spline
$A - F_1$	3	2	1	2	–	–	2
$B - F_1$	4	1	2	2	–	1	–
$A - F_2$	3	2	1	–	–	–	4
$B - F_2$	3	2	1	–	–	–	4
$A - F_3$	4	1	2	–	1	1	1
$B - F_3$	3	2	2	2	–	–	1
$A - F_1 - \alpha_1$	2	3	1	2	–	1	2
$B - F_1 - \alpha_1$	3	2	2	1	1	–	1
$A - F_2 - \alpha_1$	2	3	2	–	–	–	2
$B - F_2 - \alpha_1$	2	3	2	–	–	–	3
$A - F_3 - \alpha_1$	2	3	2	1	–	–	2
$B - F_3 - \alpha_1$	2	3	2	1	–	–	2
$A - F_1 - \alpha_2$	2	3	1	4	–	–	–
$B - F_1 - \alpha_2$	3	2	2	1	–	–	2
$A - F_2 - \alpha_2$	2	3	2	–	1	–	2
$B - F_2 - \alpha_2$	2	3	2	–	–	–	3
$A - F_3 - \alpha_2$	5	–	2	2	–	1	–
$B - F_3 - \alpha_2$	3	2	2	2	–	–	1
Total	50	40	31	20	3	4	32

modeling case study (Section 5.3) . For all analytical test cases, two alternative training samples (A and B) are used to check the sensitivity of the proposed approach to the design of experiment. The weights are calculated using three neighbors, and the test data sets are a mesh of 10 × 10 points and 5^6 points for the analytical test cases F_1 and F_2, and F_3, respectively, and 13 selected points for the ASP modeling case study.

Table 8 Characterization of the models in the ensemble of five models based on the parameters C and ϵ for the different scenarios. The sequence $X - F_i - Y$ represents the sample, the test function and noise level (if applicable)

Training sample	ϵ-insensitive			C				
	0.00	0.05	0.10	$0.25C_{\mathrm{cm}}$	$0.50C_{\mathrm{cm}}$	$0.75C_{\mathrm{cm}}$	$1.00C_{\mathrm{cm}}$	$1.50C_{\mathrm{cm}}$
$A - F_1$	1	2	–	1	–	–	–	4
$B - F_1$	3	–	1	–	1	1	–	3
$A - F_2$	2	1	–	2	1	–	–	2
$B - F_2$	3	–	–	2	–	–	–	3
$A - F_3$	3	1	–	3	–	–	–	2
$B - F_3$	3	–	–	3	–	1	–	1
$A - F_1 - \alpha_1$	–	1	1	1	–	1	–	3
$B - F_1 - \alpha_1$	–	3	–	1	–	–	–	4
$A - F_2 - \alpha_1$	1	–	1	1	1	–	–	3
$B - F_2 - \alpha_1$	–	2	–	2	1	–	–	2
$A - F_3 - \alpha_1$	2	–	–	1	–	–	–	4
$B - F_3 - \alpha_1$	2	–	–	1	–	–	–	4
$A - F_1 - \alpha_2$	1	–	1	–	–	1	–	4
$B - F_1 - \alpha_2$	2	1	–	2	–	–	–	3
$A - F_2 - \alpha_2$	–	1	1	1	–	–	–	4
$B - F_2 - \alpha_2$	1	1	–	1	–	–	–	4
$A - F_3 - \alpha_2$	5	–	–	1	2	–	–	2
$B - F_3 - \alpha_2$	3	–	–	2	1	–	1	1
Total	32	13	5	25	7	4	1	53



(a) Test function F_1

(b) Test function F_2

Fig. 3 Models that provide the best prediction at training locations in the input space for test functions F_1 and F_2. The numbers represent the model rank based on cross-validation error

The sensitivity of the approach to the number of models in the ensemble (5, 10, 18) and to the number of v nearest neighbors (3, 5) for computing the local prediction variance is also evaluated.

5.2 Analytical test functions

The analytical test functions (F_1, F_2 and F_3) with the corresponding domains of interest are shown in (7 –9). The functions F_1 and F_2 are shown in Fig. 1. Values for the coefficients in function F_3 are shown in Table 4. Note that functions F_1 and F_2, and F_3 exhibit two and six dimensions, respectively.

$$F_1(x) = [30 + x_1 \cdot \sin(x_1)] \left\lfloor 4 + \exp(-x_2)^2 \right\rfloor \qquad (7)$$
$$0 \le x_1 \le 9 \qquad 0 \le x_2 \le 6$$

$$F_2(x) = \sin\left(\frac{\pi \cdot x_1}{12}\right) \cdot \cos\left(\frac{\pi \cdot x_2}{16}\right) \qquad (8)$$
$$-10 \le x_1 \le 10 \qquad -20 \le x_2 \le 20$$

$$F_3(x) = \sum_{i=1}^{4} c_i exp\left\{-\sum_{j=1}^{6} a_{ij}\left(x_j - p_{ij}\right)^2\right\} \quad 0 \le x_j \le 1 \qquad (9)$$

5.3 Alkali-surfactant-polymer (ASP) modeling

Previous works toward the modeling and optimization of ASP processes have concentrated mainly around identifying formulations that will achieve minimum interfacial tension using laboratory experiments and empirical correlations (Bourrel et al. 1980; Salager 1996; Salager et al. 1979a,b), and sensitivity analyses using numerical simulation at core and field scale levels (Carrero et al. 2007; Hernández et al. 2001; Manrique et al. 2000; Qi et al. 2000; Wei-Ju 1996; Zhijian et al. 1998). See Zerpa et al. (2005) for details. Formal ASP flooding analysis and optimization efforts have been very limited mainly due to the high computational cost exhibited by the numerical simulations at the reservoir level, which makes impractical the coupled execution of the simulator and optimization algorithms.

The design of an ASP flooding process must achieve three main objectives: propagation of the chemicals in an active mode, the injection of enough chemicals accounting for the retention, and a complete swept of the area of interest (Lake 1989). Achieving these objectives is significantly affected by the selection of the

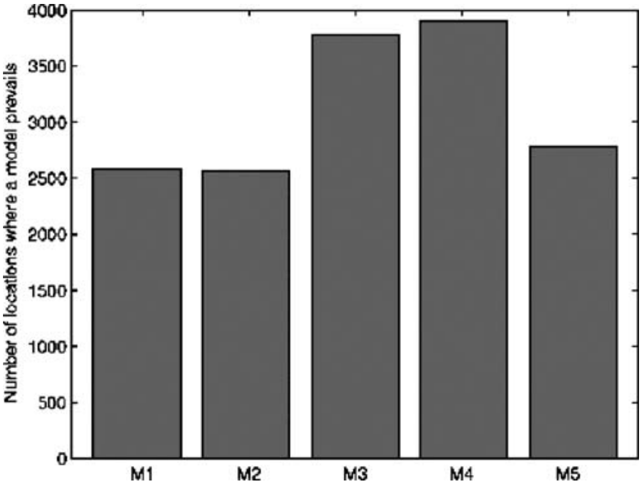
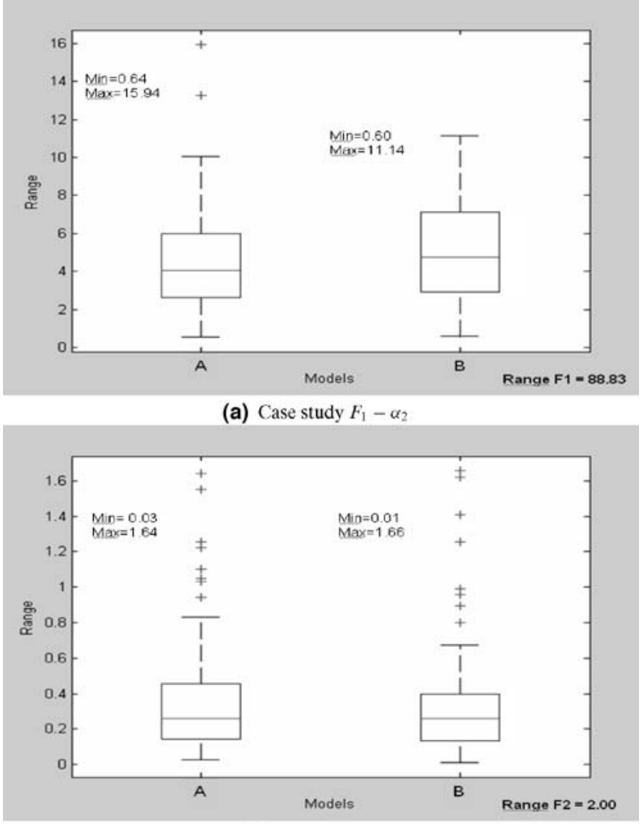


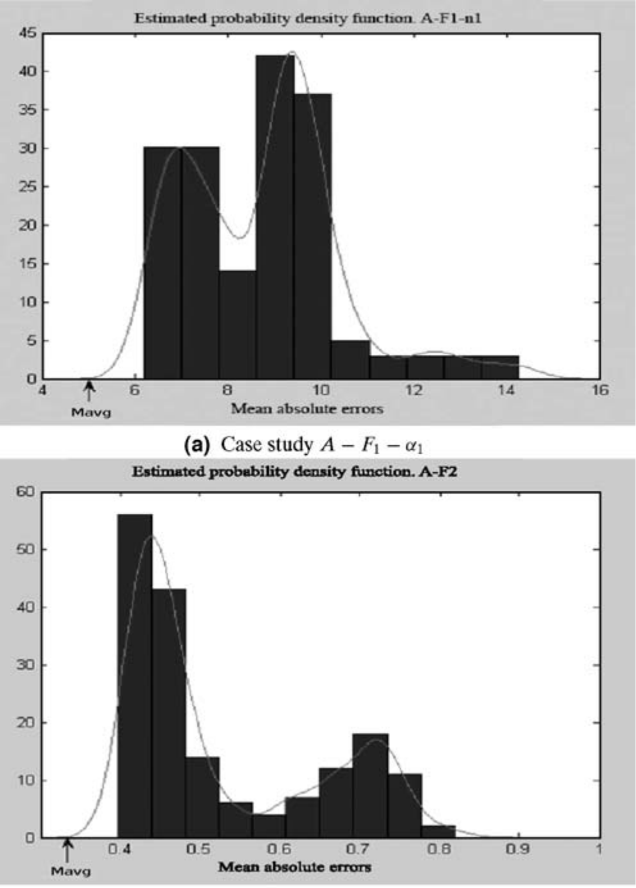
Fig. 4 Frequency of models that provide the best prediction at training locations for test function F_3. The letters Mi make reference to the model ranked i based on cross-validation error



(a) Case study $F_1 - \alpha_2$

(b) Case study F_2

Fig. 5 Empirical distribution of the errors at each training location of the models in the ensemble for selected case studies



(a) Case study $A - F_1 - \alpha_1$

(b) Case study $A - F_2$

Fig. 7 Frequency histograms of the mean absolute value of the errors at test locations for all available models and selected case studies

chemicals, the concentration of the ASP solution and the slug size, among other factors.

The ASP enhanced oil recovery modeling problem addressed here is to build a surrogate model of a computationally expensive numerical simulator that will take as input: concentration of alkaline, surfactant and polymer, and ASP slug size (expressed in the form of the injection time), and as output, the cumulative oil production. The ranges of the input variables are presented in Table 5. The cumulative oil production is calculated at 487 days. As illustrated in Fig. 2, the ASP flooding pilot has an inverted five-spot pattern

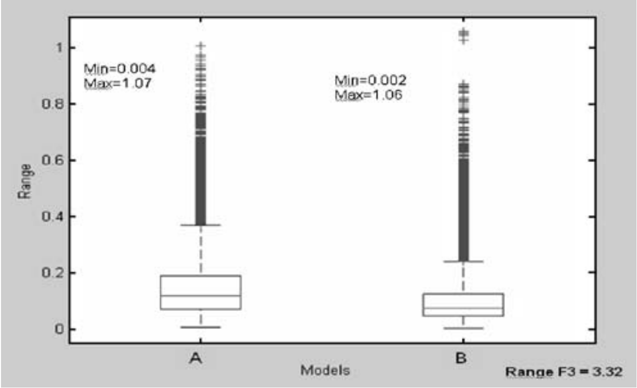


Fig. 6 Empirical distribution of the errors at each training location of the models in the ensemble for the case study $F_3 - \alpha_1$

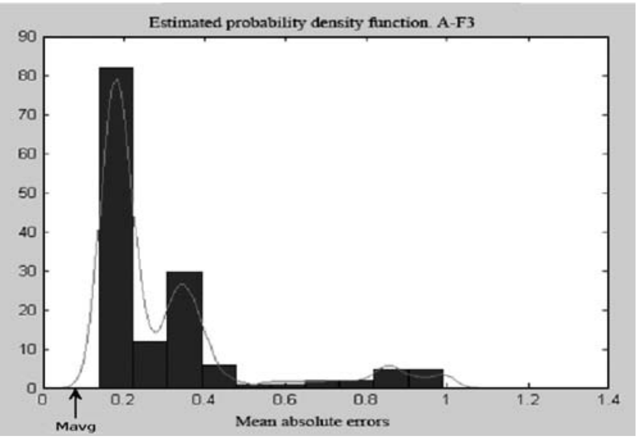
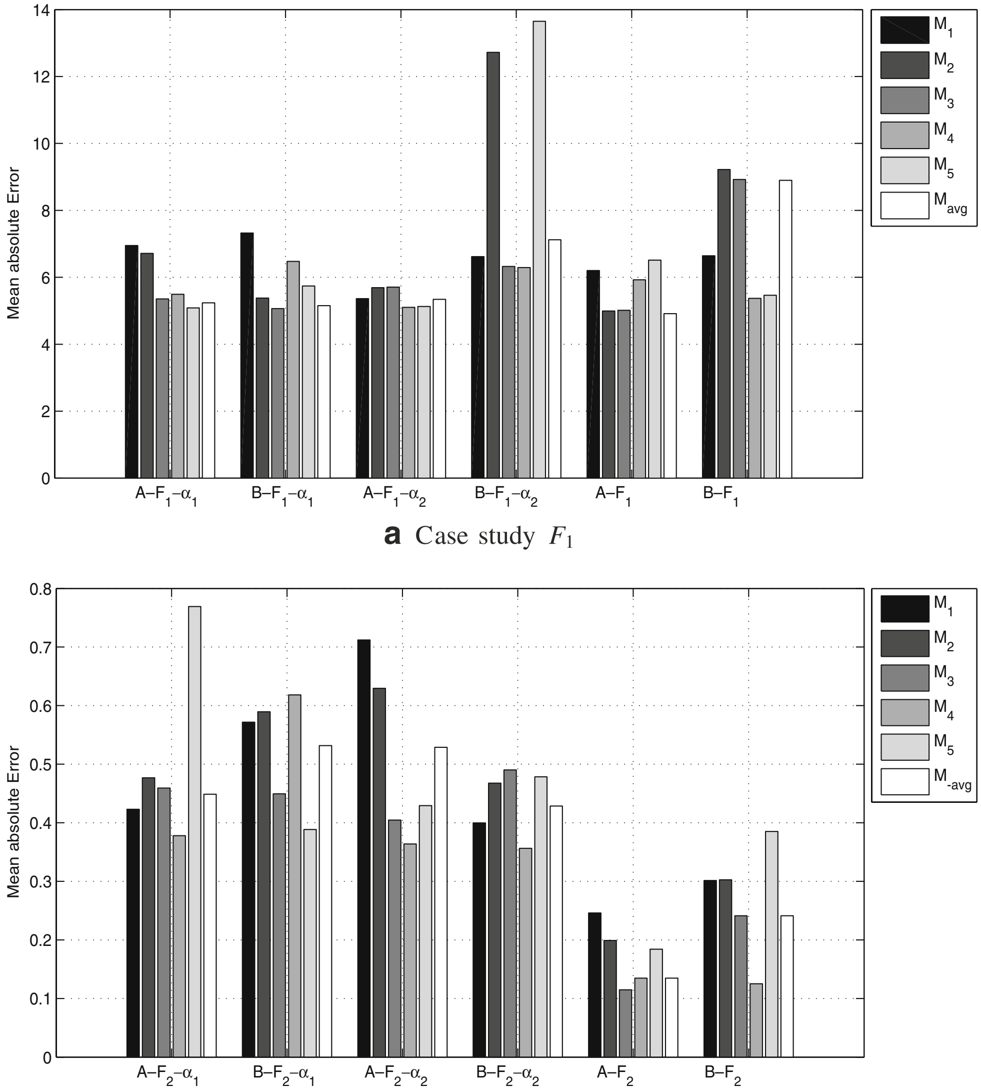


Fig. 8 Frequency histograms of the mean absolute value of the errors at test locations for all available models and case study $A - F_3$

Fig. 9 Mean absolute error of the members of the ensemble and the ensemble model for a variety of scenarios. The letter M_i makes reference to the model ranked i based on cross-validation error



a Case study F_1

b Case study F_2

Fig. 10 Mean absolute error of the members of the ensemble and the ensemble model for a variety of scenarios associated with test function F_3. The letter M_i makes reference to the model ranked i based on cross-validation error

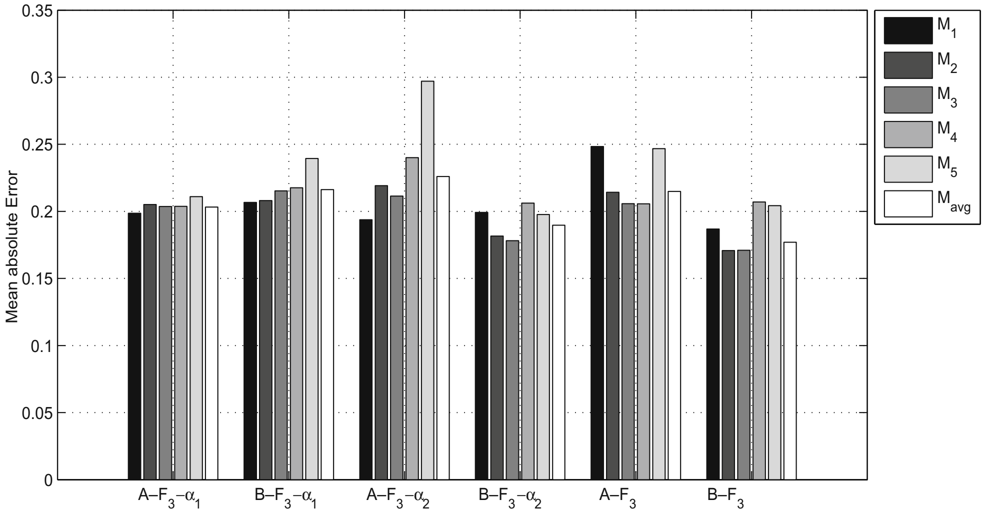


Table 9 Frequency of the rank of the ensemble model using a local performance measure (lowest error) at test locations, with respect to the individual models in the ensemble for the different scenarios

	Rank	F_1		F_2		F_3	
		Sample A	Sample B	Sample A	Sample B	Sample A	Sample B
		No. points	No. points	No. points	No. points	No. points	No. points
No noise	First	9	5	0	7	1,537	771
	Second	30	34	38	26	1,606	1,320
	Third	32	21	32	41	5,700	5,499
	Fourth	27	19	21	11	5,389	7,684
	Fifth	7	21	9	15	1,384	351
	Sixth	–	–	–	–	–	–
	Total	100	100	100	100	15,625	15,625
Noise 5% (α_1)	First	20	8	0	0	893	1,210
	Second	5	25	2	11	1,848	2,201
	Third	39	32	48	56	5,858	5,363
	Fourth	32	26	41	30	6,174	5,703
	Fifth	4	9	9	3	852	875
	Sixth	–	–	–	–	–	–
	Total	100	100	100	100	15,625	15,625
Noise 10% (α_2)	First	1	7	3	2	1,424	1,136
	Second	9	12	7	9	2131	1,462
	Third	46	37	35	47	4,860	6,648
	Fourth	38	44	49	34	6,302	6159
	Fifth	6	0	6	8	908	220
	Sixth	–	–	–	–	–	–
	Total	100	100	100	100	15,625	15,625

and a total of 13 vertical wells, nine producers and four injectors. The reservoir is at a depth of 4,150 ft, has an average initial pressure of 1,770 psi, and the porosity is assumed to be constant throughout the reservoir and equal to 0.3. The numerical grid is composed of 19 x 19 x 3 blocks in the x, y and z directions. The original oil in place is 395,427 bbls, 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. A summary of the reservoir and fluid properties is presented in Table 6. The injection scheme and other reference configuration details can be found in the sample data files of the UTCHEM program (UTCHEM-9.0 2000).

The UTCHEM program is a three-dimensional, multiphase, multicomponent reservoir simulator of chemical flooding processes developed at the University of Texas at Austin (Engelsen et al. 1987; Lake et al. 1990; Pope and Nelson 1978). 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

Table 10 Rank of the ensemble model using global error performance measures at test locations, with respect to the individual models in the ensemble for the different scenarios

	Sample	F_1			F_2			F_3		
		ma	std	max	ma	std	max	ma	std	max
No noise	A	First	First	Second	Second	Second	Third	Third	Fourth	Third
	B	Third	Fourth	Fourth	Second	Second	Fourth	Fourth	Third	Second
Noise 5% (α_1)	A	First	Second	First	Third	Third	Fourth	First	Second	Fourth
	B	Second	Second	First	Third	Third	Third	First	Fourth	Third
Noise 10% (α_2)	A	Fourth	Third	Fourth	Fourth	Fourth	Fourth	Fourth	Fourth	Fourth
	B	Second	Fourth	First	Third	Third	Third	First	Third	First

Table 11 Surrogate models error performance (ASP modeling case study)

Model	Kernel	Loss function	ma	std	max
M_1	RBF	ϵ-Insensitive	1.95	2.90	5.62
M_2	ERBF	ϵ-Insensitive	2.14	2.92	5.42
M_3	Polynomial	ϵ-Insensitive	2.33	2.97	5.90
M_4	RBF	ϵ-Insensitive	0.94	2.84	4.43
M_5	Spline	ϵ-Insensitive	1.95	3.18	5.62
M_{avg}	–	–	1.87	2.74	5.15

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 blackoil 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.

Three flowing phases and 11 components are considered in the numerical simulations. The phases are water, oil, and microemulsion, while the components are water, oil, surfactant, polymer, chloride anions, divalent cations (Ca^{2+}, Mg^{2+}), 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.

6 Results and discussion

Table 7 shows the selected models (step 4 in the solution methodology) among those in consideration for the analytical test functions F_1 ,F_2 and F_3 with and without noise. Note the diversity of the models in the ensemble for the different scenarios, with no prevailing loss function; in contrast, RBF, ERBF, and B-Spline were heavily favored as kernel functions. On the other hand, Table 8 shows the parameters C and ϵ obtained for the models in the ensemble; lower values of the parameter ϵ were frequently selected, with the parameters $0.25C_{cm}$ and $1.5C_{cm}$ associated with the best results.

Figures 3 and 4 illustrate the models that provide the best prediction throughout different regions of the input space for test functions F_1, F_2 (Fig. 3), and F_3 (Fig. 4). Note that no individual model outperforms the others, and, the range of the errors of the models at the training locations is, in general, significant. Figures 5 and 6 show the range of errors for selected test cases. Hence, in a real setting, at a particular location it is not known in advance which individual model will prevail; so an average model that weights the influence of individual models based on local measures of their error can be a more robust alternative than using any single model.

Selecting a single model from those in consideration can be risky. Figures 7 and 8 show frequency histograms of the mean absolute value of the errors for all available models and the mean absolute value of the error corresponding to the ensemble. Note the wide range of possible errors and the low value of the error associated with the ensembles when compared to the central value of the errors in the histograms. In addition, the best model selected based on training data performance was, in general, outperformed by the ensemble model, with the latter providing a more robust behavior (Figs. 9 and 10).

Tables 9 and 10 show the relative performance of the ensemble model for all scenarios with respect to the members of the ensemble using both local (Table 9) and global (Table 10) performance measures. When using local performance measures, in general, more than sixty percent of the time, the ensemble model was among the top three models using the test data, and never provided the worst performance when compared to the best individual models. On the other hand, when using mean absolute error, maximum error, and, standard deviation of the error, and considering all scenarios, the ensemble model was among the top three models sixty percent of the times, and, again,

Table 12 Frequency of the rank of the ensemble model using a local performance measure (lowest error) at test locations, with respect to the individual models in the ensemble (ASP modeling case study)

Rank	No. points
First	2
Second	0
Third	5
Fourth	5
Fifth	1
Sixth	–

Table 13 Global error performance measures for different ensemble sizes at test locations. The sequence $X - F_i$ represents the sample and test function

Sample	Ensemble size	ma	std	max
	5	4.9142	4.7090	29.2471
$A - F_1$	10	5.7348	5.7452	36.7156
	18	6.2841	5.8181	35.6998
	5	8.8954	10.0967	42.7410
$B - F_1$	10	9.2224	9.6538	40.1760
	18	9.8320	9.9998	40.6460
	5	0.1347	0.1931	0.6079
$A - F_2$	10	0.1033	0.1814	0.4066
	18	0.0943	0.1676	0.3031
	5	0.2411	0.1887	1.1950
$B - F_2$	10	0.1647	0.1301	0.9255
	18	0.1147	0.0891	0.7291
	5	0.2149	0.2299	2.0730
$A - F3$	10	0.2129	0.2284	2.0694
	18	0.2153	0.2269	2.0929
	5	0.1770	0.1976	1.9180
$B - F3$	10	0.1794	0.2000	1.8940
	18	0.1876	0.1996	1.7735

Table 14 Rank of the ensemble model with reference ensemble size with respect to those with size 10 and 18 using global error performance measures at test locations

Sample	F_1			F_2			F_3		
	ma	std	max	ma	std	max	ma	std	max
A	First	First	First	Third	Third	Third	Second	Third	Second
B	First	First	First	Third	Third	Third	First	Fist	Third

Table 15 Frequency of the rank of the ensemble model of size 10 using a local performance measure (lowest error) at test locations, with respect to the individual models in the ensemble for different scenarios

Rank	F_1		F_2		F_3	
	Sample A	Sample B	Sample A	Sample B	Sample A	Sample B
	No. points	No. points	No. points	No. points	No. points	No. points
First	8	–	3	3	956	689
Second	4	2	2	5	925	873
Third	7	3	1	3	1,012	1,071
Fourth	19	8	8	4	1,498	1,635
Fifth	22	20	21	29	2,471	4,183
Sixth	19	30	28	29	4,196	3,227
Seventh	10	30	27	30	3,002	2,884
Eighth	7	7	8	4	1,255	972
Ninth	4	–	2	2	296	84
Tenth	–	–	–	–	14	7
Eleventh	–	–	–	–	–	–
Total	100	100	100	100	15,625	15,625

Table 16 Frequency of the rank of the ensemble model of size 18 using a local performance measure (lowest error) at test locations with respect to the individual models in the ensemble for different scenarios

Rank	F_1		F_2		F_3	
	Sample A	Sample B	Sample A	Sample B	Sample A	Sample B
	No. points	No. points	No. points	No. points	No. points	No. points
First	12	–	4	1	705	866
Second	5	1	1	2	695	906
Third	3	1	4	–	739	927
Fourth	4	4	2	5	783	955
Fifth	3	4	4	7	916	1,013
Sixth	8	3	6	4	1,110	1,205
Seventh	12	6	3	6	1,338	1,535
Eighth	10	14	5	7	1,568	1,764
Ninth	20	25	6	14	1,821	1,763
Tenth	9	18	4	9	2,392	1,912
Eleventh	6	15	8	11	1,786	1,415
Twelfth	3	6	16	10	998	793
Thirteenth	–	–	17	21	567	412
Fourteenth	3	1	4	2	164	128
Fifteenth	2	0	6	1	38	31
Sixteenth	–	2	–	–	5	–
Seventeenth	–	–	–	–	–	–
Eighteenth	–	–	–	–	–	–
Nineteenth	–	–	–	–	–	–
Total	100	100	100	100	15,625	15,625

was never among the two worst models considering the best individual models. Note that the noise in the test functions did not significantly affect the ensemble performance.

Table 11 shows the ensemble model performance using the test data set for the ASP modeling case study. The ensemble model was among the top two models regardless of the performance measure (mean absolute error, standard deviation and maximum absolute error) under consideration. In addition, locally, the ensemble model was among the top three models more than half of the times and was never the worst when compared to the best individual models (Table 12).

Next, an assessment is made of the sensitivity of the results to parameters such as the number of: models in the ensemble and nearest neighbors used in the estimation of the local prediction variance. Note that the latter parameter is used to calculate the weights assigned to the models in the ensemble.

Table 13 presents global error performance measures using the test data set for three ensemble model sizes, namely, 5, 10, and 18 . The results exhibit error performance measures that can vary with the ensemble size, but no general trend was observed; that is, depending on the test function, error measure, and design of experiment, different ensemble sizes showed the lowest

Table 17 Global error performance measures for different values of the number of nearest neighbor parameter using the reference ensemble model

Case study	$v = 3$			$v = 5$		
	ma	std	max	ma	std	max
$A - F_1$	4.9142	4.7090	29.2471	4.8326	4.7335	29.8100
$B - F_1$	8.8954	10.0967	42.7410	8.6466	10.1471	42.1527
$A - F_2$	0.1347	0.1931	0.6079	0.1317	0.1952	0.6016
$B - F_2$	0.2411	0.1887	1.1950	0.2577	0.2162	1.3987
$A - F_3$	0.2149	0.2602	2.0730	0.2145	0.2597	2.0796
$B - F_3$	0.1770	0.1998	1.9180	1.1705	0.1987	1.9190

The sequence $X - F_i$ represents the sample and test function.

error. Table 14 shows the relative position of the ensemble with size 5 with respect to those of sizes 10 and 18 which confirm the previous observation. Nevertheless, in general, all the ensembles gave reasonable approximations to the test functions. When using a local performance measures for all test functions and design of experiments for ensembles of sizes 10 (Table 15) and 18 (Table 16), the ensemble model often outperformed the best individual models and was never the worst among the best individual models. In summary, while the ensemble size affects the ensemble performance, the sensitivity, after the reference ensemble size, is not strong enough to significantly deteriorate the robustness of the proposed approach. Similarly, the effect of increasing the nearest neighbors number from the reference value to 5 did not have an impact on global error performance measures such as those shown in Table 17. On more general settings, a sensitivity study is recommended, in particular, considering that the proposed approach can be conducted without human intervention and at a reasonable computational cost.

7 Conclusions

This section provides a brief description of the proposed approach for model selection, evaluation procedures, main findings and possible extensions.

- This paper presented a general approach toward the optimal selection and ensemble (weighted average) of surrogates (kernel-based approximations) to address the issue of model selection. Kernel-based regression provides an ideal setting for generating alternative models, and building ensembles of surrogates have been shown to be a worthy alternative to model selection. The surrogates for the ensemble are chosen based on their performance, favoring non-dominated models, while the weights are adaptive and inversely proportional to estimates of the local prediction variance of the individual surrogates.
- The proposed approach was evaluated using well-known analytical test functions (in two and six dimensions) and, in the surrogate-based modeling of a field scale alkali-surfactant-polymer (ASP) enhanced oil recovery process considering quadratic and ϵ-insensitive loss functions and kernels for polynomial regression, cubic splines, cubic B-splines, Gaussian radial basis functions, and exponential radial basis functions.
- It was shown that in general, the best prediction throughout the input space is given by different surrogates, and the range of the errors of the models at the training locations is, in general, significant; hence, selecting a single model can be risky, and even the best model selected based on training data performance was frequently outperformed by the ensemble of surrogates when evaluated using test data.
- When using local performance measures, in general, more than 60% of the times, the ensemble model was among the top three models using the test data and never provided the worst performance. On the other hand, when using mean absolute error, maximum error, and, standard deviation of the error and considering all scenarios, the ensemble model was among the top three models 60% of the times, and, again, was never among the two worst models. Note that the results for the test functions with noise did not affect the ensemble performance.

The proposed ensemble approach: (1) showed to be effective within the context of both analytical and engineering case studies, (2) could be extended to automatically set the number of nearest neighbors and ensemble size to optimally perform model selection, hence providing even more specific guidelines to practitioners, and (3) holds promise to be useful in more general engineering analysis and optimization scenarios.

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

References

Balabanov VO, Haftka RT, Grossman B, Mason WH, Watson LT (1998) Multidisciplinary response model for HSCT wing bending material weight. In: 7th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization, St. Louis, MO, AIAA Paper 98-4804

Bishop C (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Bourrel M, Salager JL, Schechter RS, Wade WH (1980) A correlation for phase behavior of nonionic surfactants. J Colloid Interface Sci 75(2):451–461

Buckland ST, Burnham KP, Augustin NH (1997) Model selection: an integral part of inference. Biometrics 53:275–290

Carrero E, Queipo N, Pintos S, Zerpa L (2007) Global sensibility analysis of asp enhanced oil recovery process. J Pet Sci Eng doi:10.1016/j.petrol.2006.11.007

Cherkassky V, Ma Y (2003) Comparison of model selection for regression. Neural Comput 15:1691–1714

Cherkassky V, Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw 17(1):113–126

Cherkassky V, Shao X, Mulier F, Vapnik V (1999) Model complexity control for regression using VC generalization bounds. IEEE Trans Neural Netw 10(5):1075–1089

Clarke SM, Griebsch JH, Simpson TW (2005) Analysis of support vector regression for approximation of complex engineering analyses. ASME J Mech Des 127(6):1077–1087

Craig KJ, Stander N, Dooge DA, Varadappa S (2002) MDO of automotive vehicles for crashworthiness using response surface methods. In: 9th AIAA/ISSMO symposium on multidisciplinary analysis and optimization, Atlanta, GA, AIAA Paper 2002-5607

Dosher TM, Wise FA (1976) Enhanced oil recovery potential: an estimate. J Pet Technol, p 575, Paper SPE 5800

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 Reserv Eng 427–432

Girosi F (1998) An equivalence between sparse approximation and support vector machines. Neural Comput 10(6):1455–1480

Giunta AA, Balabanov V, Grossman B, Burgee S, Haftka RT, Mason WH, Watson LT (1997) Multidisciplinary optimization of a supersonic transport using design of experiments theory and response surface modelling. Aeronaut J 101(1008):347–356

Goel T, Haftka RT, Shyy W, Queipo NV (2007) Ensemble of surrogates. Struct Multidisc Optim 33(3):199–216

Gunn SR (1998) Matlab support vector machine toolbox. http://www.isis.ecs.soton.ac.uk/isystems/kernel (Mar)

Hernández C, Chacón L, Anselmi L, Baldonedo A, Qi J, Phillip C, Pitts MJ (2001) ASP system design for an offshore application in the La Salina Field, Lake Maracaibo. In: SPE latin american and caribbean petroleum engineering conference, Buenos Aires, Argentina, Paper SPE 69544, March 2001

Hoeting J, Madigan D, Raftery A, Volinsky CT (1999) Bayesian model averaging: a tutorial. Stat Sci 14(4):382–417

Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidisc Optim 23(1):1–13

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795

Krogh A, Sollich P (1997) Statistical Mechanics of Ensemble Learning. Phys Rev E 55(1):811–825

Kurtaran H, Eskandarian A, Marzougui D, Bedewi NE (2002) Crashworthiness design optimization using successive response surface approximations. Comput Mech 29: 409–421

Lake LW (1989) Enhanced oil recovery. Prentice Hall, Englewood Cliffs, NJ

Lake LW, Bhuyan D, Pope GA (1990) Mathematical modelling of high-ph chemical flooding. SPE Paper 17398, SPE Reserv Eng 213–220

Li W, Padula S (2004) Approximation methods for conceptual design of complex systems. In: Schumaker L, Chui C, Neaumtu M (eds) Eleventh international conference on approximation theory, May 2004

Madigan D, Raftery AE (1994) Model selection and accounting for model uncertainty in graphic models using Occam's window. J Am Stat Assoc 89:1535–1546

Manrique E, De Carvajal G, Anselmi L, Romero C, Chacón L (2000) Alkali/Surfactant/Polymer at VLA 6/9/21 field in Maracaibo Lake: experimental results and pilot project design. In: SPE/DOE improved oil recovery symposium, Tulsa, OK, Paper SPE 59363, April 2000

Martin JD, Simpson TW (2005) On the use of kriging models to approximate deterministic computer models. AIAA J 43(4):853–863

Müller K, Mika S, Rätsch G, Tsuda K, Schölkopf B (2001) An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw 12(2):181–201

Perrone M (1994) General averaging results for convex optimization. In: Mozer MC et al. (ed) Proceedings of the 1993 connectionist models summer school. Erlbaum, Hillsdale, NJ

Perrone M, Cooper L (1993) When networks disagree: ensemble methods for hybrid neural networks. In: Mammone RJ (ed) Artificial neural networks for speech and vision. Chapman & Hall, FL, pp 126–142

Poggio T, Smale S (2003) The mathematics of learning: dealing with data. Not Am Math Soc 50(5):537–544

Pope GA, Nelson RC (1978) A chemical flooding compositional simulator. Paper SPE 6725, Soc Pet Eng J, p 18

Qi Q, Hongjun G, Dongwen L, Ling D (2000) The pilot test of ASP combination flooding in Karamay oil field. In: SPE international oil and gas conference an exhibition, Beijing, China, Paper SPE 64726, November 2000

Queipo NV, Haftka R, Shyy W, Goel T, Vaidyanathan R, Kevin Tucker P (2005) Surrogate-based analysis and optimization. J Progress Aerospace Sci 41:1–28

Queipo NV, Goicochea J, Pintos S (2002) Surrogate modeling-based optimization of SAGD processes. J Pet Sci Eng 35 (1–2):83–93

Queipo NV, Verde A, Canelón J, Pintos S (2002) Efficient global optimization of hydraulic fracturing designs. J Pet Sci Eng 35(3–4):151–166

Salager JL (1996) Quantifying the concept of physico-chemical formulation in surfactant–oil–water systems-state of the art. Progr Colloid Polym Sci 100:137–142

Salager JL, Bourrel M, Schechter RS, Wade WH (1979a) Mixing rules for optimum phase behavior formulation of surfactant–oil–water systems. Soc Pet Eng J 19:271–278

Salager JL, Morgan J, Schechter RS, Wade WH, Vasquez E (1979b) Optimum formulation of surfactant-oil-water systems for minimum tension and phase behavior. Soc Pet Eng J 19:107–115

Schölkopf B, Smola AJ (2002) Learning with kernels. MIT Press, Cambridge, MA

Simpson TW, Peplinski JD, Koch PN, Allen JK (2001) Metamodels for computer based engineering design: survey and recommendations. Eng Comput 17(2):129–150

UTCHEM-9.0. (2000) Utchem-9.0 a three-dimensional chemical flood simulator. http://www.cpge.utexas.edu/utchem/ (Jul)

Vapnik V (1998) Statistical learning theory. Wiley, New York

Wahba G (2000) An introduction to model building with reproducing kernel Hilbert spaces. Technical Report 1020, University of Wisconsin-Madison Statistics Department (Apr)

Wei-Ju W (1996) Optimum design of field-scale chemical flooding using reservoir simulation. PhD thesis, The University of Texas at Austin, TX

Zerpa L, Queipo NV, 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

Zhijian Q, Yigen Z, Xiansong Z, Jialin D (1998) A successful ASP flooding in gudong oil field. In: SPE/DOE Improved Oil Recovery Symposium, Tulsa, OK, Paper SPE 39613, April 1998