Toward an Optimal Ensemble of Kernel-based Approximations with Engineering Applications

Egar Sanchez, Salvador Pintos, and Nestor V. Queipo

Abstract— **This paper presents a general approach toward the optimal selection and ensemble (weighted average) of surrogates (kernel-based approximations) to address the issue of model uncertainty (model selection); that is, depending on the problem under consideration and loss function (i.e., quadratic, Laplace, ϵ-insensitive) a particular modeling scheme (e.g., polynomial regression, 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. 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 (ASP) enhanced oil recovery process, the ensemble of surrogates, in general, outperformed (i.e., mean error, standard deviation, and maximum absolute error) the best individual surrogate and provided among the best predictions throughout the domains of interest.**

I. 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 [1]-[4], automotive [5], [6], and oil industries [7], [8]. 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 uncertainty (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 [9], [10]. 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 [11], [12] provide the flexibility of generating models under alternative loss functions and, in particular, support vector regression (SVR) developed by Vapnik and co-workers [13] at AT&T Labs in the mid 90s, 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. The power of SVR resides in several fronts, such as: i) 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 ii) 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 (RKHS-reproducing kernel Hilbert space) [14]. 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 multiple-surrogate based analysis and optimization.

Previous efforts in the area of model selection have focused on either: i) select a particular surrogate from a set of candidates using, for example, Akaike information criterion (AIC) [15], Bayesian information criterion (BIC) [16], [17] or cross validation methods, or, novel techniques based on learning theoretic performance bounds, such as the structural risk minimization method [18], [19], or, ii) build an ensemble of the available surrogates (weighted average) with weights calculated based on global [20]-[23] (e.g., AIC, BIC, MSE) or local [24] performance measures. The ensemble of surrogates approach accounts for model uncertainty, and there is evidence that it can provide better average predictive ability than using any single model (e.g. [25]), 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. [24] 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 toward the optimal selection and ensemble (weighted average) of kernel-based models under alternative loss functions, with weights based on empirically estimated prediction variances and evaluate its performance using both, well-known analytical test functions, and, in the surrogate-based modeling of a field

This work was supported in part by the Fondo Nacional de Ciencia, Tecnología e Innovacion (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.

The authors are with the Applied Computing Institute, Faculty of Engineering, University of Zulia, Maracaibo, 4011 Venezuela; emails: {esanchez, spintos, nqueipo}@ica.luz.ve. N.Q. corresponding author (phone: 58-261-7598630).

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 [26], [27].

II. Problem definition

Given a training sample $E = \{(x_j, y_j) : 1 \leq j \leq n\}$ of a function $y = f(x)$ defined in $D \subset R^p$, and k kernel-based surrogate models M_i, $1 \leq i \leq k$ constructed from sample E, select a set of m surrogate models and build a weighted average model:

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

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

III. 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 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 models a set of parameters are specified, namely, C (regularization parameter), ϵ (size of insensitive zone), for SVR models, the width h for Gaussian and the degree p of the polynomial kernel (see next section for details). The parameters were specified 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.
4) A model that do not dominate (lowest absolute error) the others in one or more of the training locations is discarded, unless it is one of the best three models based on the mean absolute error criterion. Using this procedure, the subset of models (m) to be part of the ensemble for prediction is created.
5) Using the subsets of 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 prediction variance for each of the models is estimated empirically using v nearest neighbors of point x. Specifically,

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

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:

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

TABLE I
Kernel functions

Kernel	Parametrization
Polynomial order d	$k(x,x') = (\langle x,x' \rangle + c)^d \quad d \in N, c \geq 0$
Cubic Spline	$k(x,x') = 1 + \langle x,x' \rangle + 1/2 \langle x,x' \rangle \min(x,x') - 1/6 \min(x,x')^3$
B Spline order $2n+1$	$k(x,x') = B_{2n+1}(\|x-x'\|)$ $B_k = \otimes_{i=1}^{k+1} I_{[-1/2,1/2]}$
Gaussian	$k(x,x') = exp\left(-\|x-x'\|^2 / 2h^2\right) \quad h > 0$

IV. Kernel-based regression

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

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

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 second term penalizes hypothesis that are too complex.

If we restrict ourselves to Reproducing Kernel Hilbert Spaces (RKHS) the variational problem can be formulated as stated in Equation (1).

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

It can be shown that independently of the form of the loss function, the solution of the variational problem (Equation (1)) can be expressed as:

$$M(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i) \quad (2)$$

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

In particular, if the loss function is quadratic the coefficients in Equation (2) can be found by solving the following linear system:

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

TABLE II

SURROGATE MODELS UNDER CONSIDERATION (CASE STUDIES)

Models	Loss function		Kernels			
	ϵ-insensitive	Quadratic	Gaussian	Polynomial	Spline	Bspline
1	X		X			
2	X			X		
3	X				X	
4	X					X
5		X	X			
6		X		X		
7		X			X	
8		X				X

where K denotes the so called Gram matrix with component K_{ij} denoting $k(x_i, x_j)$, and I representing the identity matrix. Alternatively, if the ϵ-insensitive loss function is used, the coefficients in Equation (2) are found by solving a quadratic programming problem. See Schï¿½kopf and Smola [28], and Poggio and Smales [29] for details.

V. CASE STUDIES

A. General considerations

The solution methodology is evaluated using two well-known [10] test functions (F_1 and F_2) and a modeling problem in the area of enhanced oil recovery. Table II shows the models under consideration (quadratic and ϵ-insensitive loss functions; and polynomial, Gaussian radial basis functions, splines and B-splines). Third and second order polynomials, and third degree splines were considered. The kernel-based regression problems were solved using the Matlab Support Vector Machines (SVM) toolbox [30]. The ϵ and C values under consideration were 0, 0.05, 0.1 and $0.5\,C_{cm}$, $0.75\,C_{cm}$, $1.00\,C_{cm}$, $1.5\,C_{cm}$, respectively; the h values were set equal to $1.00h_{cm}$ and $0.50h_{cm}$, with C_{cm} and h_{cm} reference values as proposed by Cherkassky and Ma [31]. The parameter values were selected using cross validation (k-fold) using twenty 20 training points mesh with $k = 5$ for the analytical test cases (Section V-B), and sixty four (64) training points in a latin hypercube sample with $k = 8$ for the ASP modeling (Section V-C) case study. The weights in all cases were calculated using three (3) neighbors. The test data sets were a mesh of 10×10 points for the analytical test cases and eight (8) selected points for the ASP modeling case study.

B. Analytical test functions

The analytical test functions (F_1 and F_2) with the corresponding domain of interest are shown in Equations (3) and (4). The functions are shown in Figures 1 and 2 and Table III display their main characteristics; namely, their dimension, the x and y location at which the minimum occurs, and range.

$$F_1(x) = [30 + x_1 \cdot \sin(x_1)]\,[4 + \exp(-x_2)^2]\,, \quad 0 \le x_1 \le 9 \quad 0 \le x_2 \le 6 \tag{3}$$

$$F_2(x) = \sin\left(\tfrac{\pi \cdot x_1}{12}\right) \cdot \cos\left(\tfrac{\pi \cdot x_2}{16}\right), \quad -10 \le x_1 \le 10 \quad -20 \le x_2 \le 20 \tag{4}$$

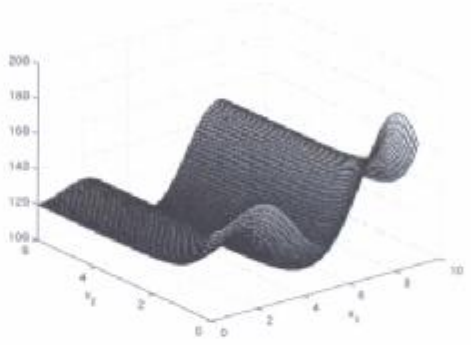

Fig. 1. F_1 test function

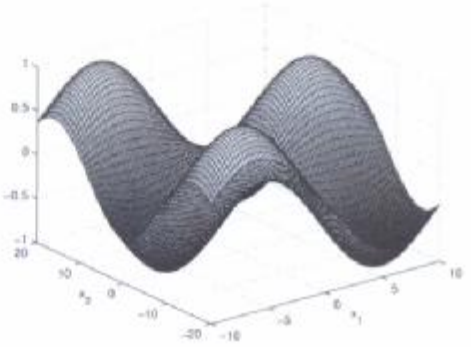

Fig. 2. F_2 test function

C. 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 [32]-[35], and sensitivity analyses using numerical simulation at core and field scale levels [36]-[40]. See Zerpa et al. [24] 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

TABLE III

TEST FUNCTIONS MAIN CHARACTERISTICS

Name	x_{opt}	y_{opt}	Range
F_1	4.9074 5.9959	100.7426	88.8295
F_2	$\begin{matrix} -6 & 0 \\ 6 & -16 \\ 6 & 16 \end{matrix}$	$\begin{matrix} -1 \\ -1 \\ -1 \end{matrix}$	2

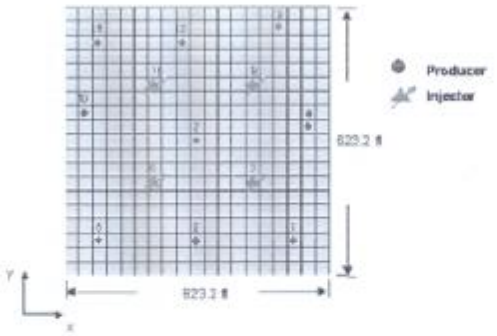

Fig. 3. Reference configuration (ASP modeling case study)

TABLE IV
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

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 [27]. Achieving these objectives is significantly affected by the selection of the 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 IV. The cumulative oil production is calculated at 487 days. As illustrated in Fig. 3, the ASP flooding pilot has an inverted five-spot pattern and a total of 13 vertical wells, 9 producers and 4 injectors. The reservoir is at a depth of 4150 ft., has an average initial pressure of 1770 psi, and the porosity is assumed to be constant throughout the reservoir and equal to 0.3. The numerical grid is composed of 19x19x3 blocks in the x, y and z directions. The OOIP 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 the Table V. The injection scheme and other reference configuration details can be found in the sample data files of the UTCHEM program [41]. Three flowing phases and eleven components are considered in the numerical

TABLE V
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	CNa	0.0583	meq/ml
	CCa	0.0025	meq/ml

TABLE VI
SELECTED SURROGATE MODELS (CASE STUDY - F_1 & F_2)

Model	F_1		F_2	
	Kernel	Loss function	Kernel	Loss function
M_1	Gaussian	ϵ-insensitive	Gaussian	ϵ-insensitive
M_2	Gaussian	Quadratic	Gaussian	Quadratic
M_3	Spline	ϵ-insensitive	Cubic-polynomial	ϵ-insensitive
M_4	Spline	Quadratic	B-Spline	Quadratic

simulations. The phases are water, oil and microemulsion, while the components are water, oil, surfactant, polymer, chloride anions, divalent cations (Ca2+, Mg2+), 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.

VI. RESULTS AND DISCUSSION

Table VI shows the selected models (step 4 in the solution methodology) among those in consideration in the case study corresponding to the analytical test functions F_1 y F_2. With reference to function F_1, Figure 4 illustrates the model that provides the best prediction throughout different regions of the input space; note that no individual model outperform the others. 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.

Table VII shows the performance of the individual models and the W_{avg} on the test data set. Figure 5 shows a box-plot with the empirical distribution of the absolute value of the errors in the test data set for all models under consideration. The W_{avg} outperformed M_1, M_2, and M_3 and have a similar performance than M_4 for all error measures (i.e., mean absolute error, standard deviation, and maximum absolute error). Specifically, the W_{avg} model outperformed the mean value of the error measures of M_1, M_2, M_3, and M_4 by at least 23%, 19%, and 6%, respectively.

With reference to function F_2, the W_{avg} outperformed the individual models under all performance measures, specifically, it gave at least 61%, 64%, and 72% lower values

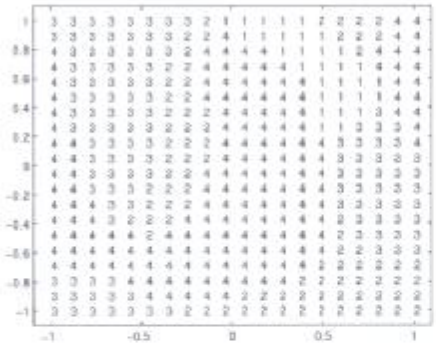

Fig. 4. Models providing the best predictions at test locations (Case study - F_1). $M_1 = 1$, $M_2 = 2$, $M_3 = 3$, $M_4 = 4$

TABLE VII

SURROGATE MODELS PERFORMANCE (CASE STUDY - F_1 & F_2)

	F_1			F_2		
Model	*dm*	*std*	*max*	*dm*	*std*	*max*
M_1	6.09	8.35	38.11	0.13	0.17	0.49
M_2	6.55	8.64	39.04	0.16	0.19	0.41
M_3	8.73	11.18	32.40	0.41	0.69	3.53
M_4	4.85	6.64	23.98	0.12	0.18	0.61
W_{avg}	5.04	7.01	31.53	0.08	0.11	0.35

of mean absolute error, standard deviation, and maximum absolute error, respectively, than the mean value of the error measures of all the individual models, and represents a 33%, 35%, and 15% improvement in the mean absolute error, standard deviation, and maximum absolute error, with respect to the corresponding error measure in the best individual model. Figure 6 shows the empirical distribution of the absolute value of the errors in the test data set.

Table VIII presents for test functions F_1 and F_2, the rank of the W_{avg} model in estimating the function values for the test data set. In general, the W_{avg} provides among the best predictions and as previously stated represents a robust estimator.

Table IX shows the selected models for the ASP modeling case study and their relative performance. The W_{avg} model outperformed the individual models under all error measures; specifically, it provided 58%, 55%, and 57% lower values

TABLE VIII

W_{avg} RANK AT TEST LOCATIONS (CASE STUDY - F_1 & F_2)

	Number of points	
Rank	F_1	F_2
1	12	21
2	36	50
3	34	28
4	18	1
5	0	0

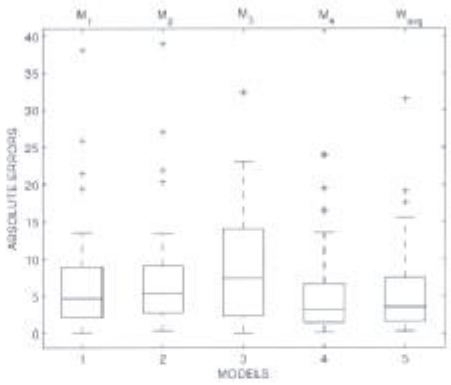

Fig. 5. Absolute errors empirical distributions (Case study - F_1)

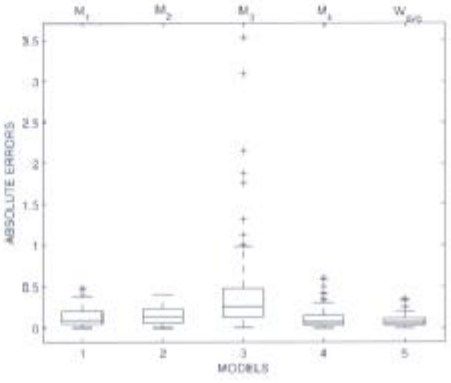

Fig. 6. Absolute errors empirical distributions (Case study - F_2)

than the mean of the error measures of all individual models, respectively, than the mean response of all models under consideration, and represents a 41%, 53%, and 22% improvement in the mean absolute error, standard deviation, and maximum absolute error, with respect to the corresponding error measure associated with the best individual model.

VII. CONCLUSIONS

- 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 uncertainty (model selection). The surrogates

TABLE IX

SURROGATE MODELS PERFORMANCE (ASP MODELING CASE STUDY)

Model	Kernel	Loss function	*dm*	*std*	*max*
M_1	Gaussian	ϵ-insensitive	1.60	2.75	7.08
M_2	Gaussian	Quadratic	2.45	3.06	6.98
M_3	Spline	ϵ-insensitive	2.66	2.71	3.18
W_{avg}			0.95	1.28	2.47

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 (F_1, F_2) and, in the surrogate-based modeling of a field scale alkali-surfactant-polymer (ASP) enhanced oil recovery process considering an initial set of eight (8) surrogate models which were the result of quadratic and ϵ-insensitive loss functions and kernels for polynomial regression, cubic splines, cubic B-splines and Gaussian radial basis functions.
- The surrogate selection strategy led to four (4) and three (3) surrogate ensembles, for the analytical, and ASP modeling case studies, and it was shown that, in general, the best prediction throughout the input space is given by different surrogates, and, the weighted average model (ensemble) outperformed (i.e., mean error, standard deviation, and maximum absolute error): i) the corresponding mean value of the individual surrogates (F_1: 23%, 19%, and 6%; F_2: 61%, 64%, and 72%; ASP modeling: 58%, 55%, and 57%) and, ii) the best individual surrogates (F_2: 33%, 35%, and 15%; ASP modeling: 41%, 53%, and 22%). In addition, the ensemble prediction typically ranked among the best when compared to those provided by the individual models.
- The proposed ensemble approach showed to be effective within the context of both analytical and engineering case studies and holds promise to be useful in more general engineering analysis and optimization scenarios.

ACKNOWLEDGMENT

The authors thank the Center for Petroleum and Geosystems Engineering of The University of Texas at Austin for providing the UTCHEM compositional simulator.

REFERENCES

[1] A. A. Giunta, V. Balabanov, S. Burgee, B. Grossman, R. T. Haftka, W. H. Mason and L. T. Watson, "Multidisciplinary optimisation of a supersonic transport using design of experiments theory and response surface modelling," *Aeronautical J.*, vol. 101, no. 1008, pp. 347-356, 1997.

[2] V. O. Balabanov, R. T. Haftka, B. Grossman, W. H. Mason and L. T. Watson, "Multidisciplinary response model for HSCT wing bending material weight," AIAA Paper 98-4804, *7th AIAA/USAF/NASA/ISSMO Symp. On Multidisciplinary Anal. and Optim.*, St. Louis, MO, 1998.

[3] W. Li and S. Padula, "Approximation methods for conceptual design of complex systems," in *Eleventh International Conference on Approximation Theory*, Edited by Chui C, Neaumtu M, Schumaker L, May 2004.

[4] N. Queipo, R. Haftka, W. Shyy, T. Goel, R. Vaidyanathan and P. Kevin Tucker, "Surrogate-based analysis and optimization," *Journal of Progress in Aerospace Sciences*, vol. 41, pp. 1-28, 2005.

[5] K. J. Craig, N. Stander, D. A. Dooge and S. Varadappa, "MDO of automotive vehicles for crashworthiness using response surface methods," AIAA paper 2002-5607, *9th AIAA/ISSMO Symp. on Multidisciplinary Anal. and Optim.*, Atlanta, GA. 2002.

[6] H. Kurtaran, A. Eskandarian, D. Marzougui and N. E. Bedewi, "Crashworthiness design optimization using successive response surface approximations," *Computational Mechanics*, vol. 29, pp. 409-421, 2002.

[7] N. V. Queipo, A. Verde, J. Canelli¿½ and S. Pintos, "Efficient global optimization of hydraulic fracturing designs," *Journal of Petroleum Science and Engineering*, vol. 35, no. 3-4, pp. 151- 166, August 2002.

[8] N. V. Queipo, J. Goicochea and S. Pintos, "Surrogate modeling-based optimization of SAGD processes," *Journal of Petroleum Science and Engineering*, vol. 35, no. 1-2, pp. 83-93, July 2002.

[9] T. W. Simpson, J. D. Peplinski, P. N. Koch and J. K. Allen, "Metamodels for computer based engineering design: Survey and recommendations," *Engineering with Computers*, 17, pp. 129-150, 2001.

[10] R. Jin, W. Chen and T. W. Simpson, "Comparative studies of metamodeling techniques under multiple modeling criteria," *Structural and Multi-disciplinary Optimization*, 23, pp. 1-13, 2001.

[11] F. Girosi, "An equivalence between sparse approximation and support vector machines," *Neural Computation*, vol. 10, no. 6, pp. 1455-1480, 1998.

[12] K. Miller, S. Mika, G. Ri¿½sch, K. Tsuda and B. Schi¿½lkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181-201, May 2001.

[13] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.

[14] G. Wahba, "An introduction to model building with reproducing kernel Hilbert spaces," *Technical Report No. 1020*, April 18, 2000.

[15] S. T. Buckland, K. P. Burnham and N. H. Augustin, "Model selection: an integral part of inference," *Biometrics*, 53, pp. 275-290, 1997.

[16] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Statist. Assoc*, vol. 90, pp. 773-795, 1995.

[17] J. Hoeting, D. Madigan, A. Raftery and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Statistical Science*, vol. 14, no. 4, pp. 382-417, 1999.

[18] V. Cherkassky, X. Shao, F. Mulier and V. Vapnik, "Model complexity control for regression using VC generalization bounds," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1075-1089, 1999.

[19] V. Cherkassky and Y. Ma, "Comparison of model selection for regression," *Neural computation*, 15, pp. 1691-1714, 2003.

[20] M. Perrone and L. Cooper, "When networks disagree: ensemble methods for hybrid neural networks," in *Artificial Neural Networks for Speech and Vision*, Edited by R. J. Mammone, London: Chapman & Hall, pp. 126-142, 1993.

[21] M. Perrone, "General Averaging Results for Convex Optimization," in *Proceedings Connectionist Models Summer School*, Edited by M. C. Mozer et al., Hillsdale, NJ: Lawrence Erlbaum, pp. 364-371, 1994.

[22] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford: University Press, pp. 364-369, 1995.

[23] T. Goel, R. T. Haftka, W. Shyy and N. V. Queipo, "Ensemble of surrogates," submitted to *Structural and Multidisciplinary Optimization Journal*.

[24] L. Zerpa, N. V. Queipo, S. Pintos and J. Salager, "An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates," *Journal of Petroleum Science and Engineering*, 47, pp. 197-208, 2005.

[25] D. Madigan and A. E. Raftery, "Model selection and accounting for model uncertainty in graphic models using Occam's window," *J. Amer. Statist. Assoc.*, vol. 89, pp. 1535-1546, 1994.

[26] T. M. Doscher and F. A. Wise, "Enhanced oil recovery potential. An estimate," Paper SPE 5800, *J. Petroleum Technology*, p. 575, May 1976.

[27] L. W. Lake, *Enhanced Oil Recovery*, Englewood Cliffs, NJ: Prentice Hall, p. 408, 1989.

[28] B. Schi¿½lkopf and A. J. Smola, *Learning with Kernels*, Cambridge, MA: MIT Press, 2002.

[29] T. Poggio and S. Smale, "The mathematics of learning: Dealing with data," *Notices of the AMS*, vol. 50, no. 5, pp. 537-544, May 2003.

[30] S. R. Gunn, "MATLAB Support Vector Machine Toolbox," http://www.isis.ecs.soton.ac.uk/isystems/kernel, March 1998.

[31] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Networks*, vol. 17, no. 1, pp. 113-126, 2004.

[32] J. L. Salager, J. Morgan, R. S. Schechter, W. H. Wade and E. Vasquez, "Optimum formulation of surfactant-oil-water systems for minimum tension and phase behavior," *Soc. Petroleum Eng. J.*, 19, pp. 107-115, 1979.

[33] J. L. Salager, M. Bourrel, R. S. Schechter and W. H. Wade, "Mixing rules for optimum phase behavior formulation of surfactant-oil-water systems," *Soc. Petrol. Eng. J.*, 19, pp. 271-278, 1979.

[34] M. Bourrel, J. L. Salager, R. S. Schechter and W. H. Wade, "A correlation for phase behavior of nonionic surfactants," *Journal of Colloid and Interface Science*, vol. 75, No. 2, pp. 451-461, June 1980.

[35] J. L. Salager, "Quantifying the concept of physico-chemical formulation in surfactant-oil-water systems-state of the art," *Progr Colloid Polym Sci*, 100, pp. 137-142, 1996.

[36] Q. Zhijian, Z. Yigen, Z. Xiansong and D. Jialin, "A Successful ASP flooding in Gudong Oil Field," paper SPE 39613, *SPE/DOE Improved Oil Recovery Symposium*, Tulsa, OK, April 1998.

[37] W. Wei-Ju, *Optimum Design of Field-Scale Chemical Flooding Using Reservoir Simulation*, Ph.D. Thesis, The University of Texas at Austin, 1996.

[38] E. Manrique, G. De Carvajal, L. Anselmi, C. Romero and L. Chacín, "Alkali/Surfactant/Polymer at VLA 6/9/21 field in Maracaibo Lake: Experimental results and pilot project design," paper SPE 59363, *SPE/DOE Improved Oil Recovery Symposium*, Tulsa, OK, April, 2000.

[39] Q. Qi, G. Hongjun, L. Dongwen and D. Ling, "The pilot test of ASP combination flooding in Karamay oil field," paper SPE 64726, *SPE International Oil and Gas Conference an Exhibition*, Beijing, China, Nov. 2000.

[40] C. Hernández, L. Chacín, L. Anselmi, A. Baldonedo, J. Qi, C. Phillip and M. J. Pitts, "ASP system design for an offshore application in the La Salina Field, Lake Maracaibo," paper SPE 69544, *SPE Latin American and Caribbean Petroleum Engineering Conference*, Buenos Aires, Argentina, March 2001.

[41] *UTCHEM-9.0, A Three-Dimensional Chemical Flood Simulator*, vol. 1 and 2, Reservoir Engineering Research Program, Center for Petroleum and Geosystems Engineering, The University of Texas at Austin, July 2000.