

A BALANCED SEQUENTIAL DESIGN STRATEGY FOR GLOBAL SURROGATE MODELING

Prashant Singh
Dirk Deschrijver
Tom Dhaene

Ghent University - iMinds
Sint Pietersnieuwstraat 41
Ghent, 9000, BELGIUM

ABSTRACT

The sequential design methodology for global surrogate modeling of complex systems consists of iteratively training the model on a growing set of samples. Sample selection is a critical step in the process and influences the final quality of the model. It is desirable to use as few samples as possible while building an accurate model using insight gained in previous iterations. A robust sampling scheme is considered that employs Monte Carlo Voronoi tessellations for exploration, linear gradients for exploitation and different schemes are investigated to balance their trade-off. The experimental results on benchmark examples indicate that some schemes can result in a substantially smaller model error especially when the system under consideration has a highly non-linear behavior.

1 INTRODUCTION

Computer experiments or simulations are extensively used in various disciplines such as medicine, environmental sciences and physical sciences, including complex system design problems such as aircraft design. In such cases, the computer models constructed to simulate experiments can be very complex and computationally intensive. The scale of computation is reflected in an example given in (Wang and Shan 2007) where it is reported that it takes the Ford Motor Company anywhere from 36 to 160 hours to run one crash simulation. For an optimization problem, the total computation time taken by repeated iterations may sum up to a practically unacceptable range of 75 days to 11 months. Valuable time, money and resources can be saved if a cheaper replacement or approximation is fit to the underlying model. This “model of the model” is called a metamodel or a surrogate model.

Surrogate models may cover the entire domain and can be used as full replacements of the underlying system. Such surrogates are called *global* surrogate models. Contrarily, *local* surrogate models only cover a region of the entire domain of underlying systems with the aim of finding the global optimum and are mostly used for design optimization. The focus of this work is on global surrogate modeling.

Surrogate modeling methodologies can be broadly divided into traditional design of experiments (DOE) and sequential design (or adaptive sampling (Lehmniesiek and Petrie Meyer 2002)). In traditional DOE the surrogate is trained in a single iteration on a carefully chosen set of samples. Such a set of samples which is aimed at capturing the maximum amount of information about the design (or input) space is called an experimental design. On the other hand, a typical sequential design surrogate modeling process starts with an initial experimental design which leads to a preliminary model, and new carefully chosen samples are iteratively added to enrich this model. This process continues till the model reaches sufficient accuracy, or until a predefined sampling budget is exceeded. This paper concerns only the sequential design methods.

The model itself can be of any type. In addition to polynomial functions, surrogates makes use of diverse methods such as splines, kriging, radial basis functions, neural networks, etc. (Wang and Shan 2007). Kriging, which is a stochastic model was first proposed as a metamodeling method by (Sacks et al.

1989). Combinations of the methods mentioned above have also been tried. Though it is not possible to pronounce a specific model superior to the rest, Kriging and related Gaussian processes have been intensely studied recently.

Since surrogates are trained using samples obtained by performing underlying (often expensive) simulations or experiments, it is desirable to use as few samples for training as possible while meeting specified accuracy goals. Thus, sampling algorithms or sample selection schemes become crucial to the success of a sequential design strategy. Sampling schemes might involve exploration (of the whole design space) or exploitation (of regions with high non-linearity) or a combination of both. In this paper, we propose an adaptive sampling scheme that combines exploration and exploitation and controls the contribution of the two according to the different balancing schemes.

The paper is organized as follows. Section 2 describes the related work and sequential design methods that are proposed in the literature. Section 3 explains the different balancing schemes that will be investigated, while Section 4 reports the experiments performed and results obtained. Finally, Section 5 concludes the paper and summarizes the results.

2 SEQUENTIAL DESIGN METHODS

There has been wide interest in the study of sequential design strategies for global surrogate modeling in the last decade. Space-filling exploration methods are investigated in (Crombecq, Laermans, and Dhaene 2011) and the authors conclude that pre-optimized latin hypercube designs should be preferred if the number of samples is known in advance, and threshold Monte Carlo methods otherwise. Continuing the use of Monte Carlo methods, (Crombecq et al. 2011) proposed a sequential design strategy (called LOLA-Voronoi) that utilizes local linear approximation of the objective function for exploitation and Monte Carlo Voronoi tessellation for exploration. Both components in this study are based on LOLA-Voronoi, with some modification performed on the exploitation component as explained below :

- The exploitation component uses an estimation of the function gradient to gauge the non-linearity at the sample location. The higher the gradient, the steeper the slope of the function is expected to be, which in turn implies high non-linearity. For a d dimensional input space x , the gradient ∇f of a function f at a given sample location can be defined as:

$$\nabla f = \left(\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \dots, \frac{\partial f}{\partial x^d} \right) \quad (1)$$

Since the function f is not known in advance, the gradient at all the existing sample points is estimated by fitting a linear model through the neighbouring sample points, and a new sample is randomly generated around points that have high gradient values. It should be noted that the range used for generating new samples randomly is inversely proportional to the gradient value. The intuition being, that a sample having a very high gradient will have high non-linearity (and thus, high uncertainty) around it, and thus the new sample should be located very close to it.

- To complement the exploitation component, a Voronoi tessellation based exploration component is included, which is the same as the one used in (Crombecq et al. 2011). The sample density of the entire design space is estimated using a Voronoi tessellation. The samples having a large Voronoi cell size are likely to have regions around them which are prone to under-sampling, and hence such regions are given priority during the sampling process to ensure that all regions in the design space are uniformly covered.

Since both criteria are possibly conflicting, the corresponding metrics are normalized in $[0, 1]$ and they are unified into a combined metric that keeps the ratio of exploration and exploitation components constant throughout the modeling process (although it can be varied before the process starts). This hybrid approach performed well in the experiments conducted by the authors. The algorithm has been used successfully

in various research domains such as microwave engineering (Deschrijver et al. 2011), bioelectromagnetics (Aerts et al. 2013), and electromagnetic compatibility (Deschrijver et al. 2012). In the next section, we propose alternative schemes that balance the trade-off between exploration and exploitation.

3 BALANCING SCHEMES

Although it has been established that both exploration and exploitation must be performed, the extent of their individual contribution has not yet been explored. In this paper, a modification to the LOLA-Voronoi algorithm is proposed by considering 3 balancing schemes : ε -greedy, ε -decreasing and ε -adaptive.

- **ε -greedy** : A tuning parameter $\varepsilon \in [0, 1]$ decides the proportion of sampling iterations for which all samples returned are purely exploration samples, and purely exploitation samples are returned for the remaining $1 - \varepsilon$ proportion of sampling iterations. In each iteration step, a random number α is generated according to a uniform distribution. If $\alpha < \varepsilon$, then the current sampling iteration contains purely explorative samples, and if $\alpha \geq \varepsilon$, then the current sampling iteration contains purely exploitative samples. The value of ε is a choice that is made up-front by the designer.
- **ε -decreasing** : This strategy is similar to ε -greedy, except for the fact that the value of ε decreases with proceeding sampling iterations. Thus, this scheme allows the algorithm to perform highly explorative behaviour up-front and ends up with a lot of exploitation during the final sampling iterations. Exploration will dominate in the beginning when the uncertainty of the model is high. As we explore more of the design space, the uncertainty goes down with time and the exploitation component takes over, eventually eclipsing exploration.
- **ε -adaptive** : Unlike the previous two schemes, this strategy builds a surrogate model in all successive sampling iterations. Each time an additional batch of sample points is selected, the model is updated and the contribution of the added sample points is quantified. This can be done by comparing the deviation between successive models, fluctuations in the global model response, and by investigating the evolution of the cross-validation error. Depending on the added contribution of a given sample batch, the ε value is adaptively increased or decreased over time in a dynamical way.

Fig. 1 presents a graphical visualization of the different schemes, where the *expected* number of batches is shown that are chosen according to the exploration and exploitation criteria as evolving over time.

4 EXPERIMENTS AND RESULTS

4.1 SUMO Toolbox

All experiments in this paper were performed in MATLAB and are compared to the LOLA-Voronoi implementation that is available in the SUMO toolbox (Gorissen et al. 2010). The toolbox is highly adaptable and supports many model types which makes it an ideal choice for the experiments. A typical modeling sequence in SUMO begins with an initial design (e.g. Latin hypercube) upon which an initial model is trained. The accuracy of the model is estimated using a chosen measure (e.g. error estimation using cross-validation), and if the desired accuracy has not been reached, or the allotted number of simulations for the run have not been exhausted, the process begins a new iteration with a sampling scheme selecting a new batch of samples to train the model. This can also be seen as a process of optimization of the model by tuning its parameters.

4.2 Benchmark Functions

The problems chosen for the experiments were modeling the Ackley, Bird and Branin functions. For the sake of simplicity and understanding, all three problems taken are 2-dimensional problems.

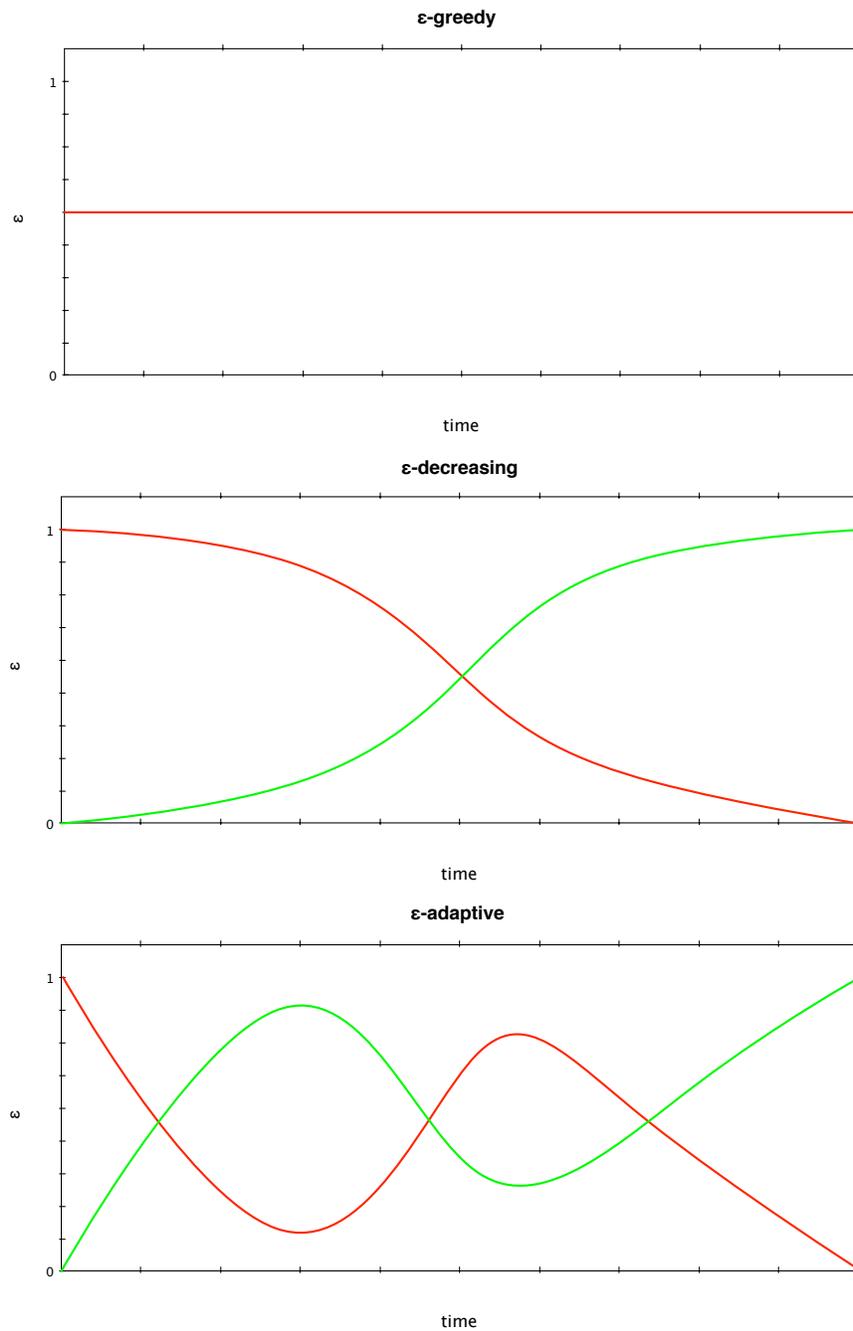


Figure 1: The figure shows the *expected* number of samples per batch (%) that are chosen over time according to exploration (red) or exploitation (green) for the ϵ -greedy, ϵ -decreasing and ϵ -adaptive algorithm.

- The Ackley function can be defined in D -dimensions and has a high number of local minima, but a single global minimum at the origin. Taking e as Euler's number, the function is defined as

$$f(x) = -20 \cdot \exp\left(-0.2 \cdot \sqrt{\frac{1}{D} \cdot \sum_{i=1}^D x_i^2}\right) - \exp\left(\frac{1}{D} \cdot \sum_{i=1}^D \cos(2\pi \cdot x_i)\right) + 20 + e \quad (2)$$

- The Bird function consists of two design variables. It has 2 global optima, and few local optima

$$f(x_1, x_2) = \sin(x_1) * \exp((1 - \cos(x_2))^2) + \cos(x_2) * \exp((1 - \sin(x_1))^2) + (x_1 - x_2)^2 \quad (3)$$

- The Branin function is also a two-variable function with $-5 \leq x_1 \leq 10$ and $0 \leq x_2 \leq 15$

$$f(x_1, x_2) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10, \quad (4)$$

4.3 Numerical Results

The initial design chosen was a Latin hypercube of 100 samples in addition to 4 corner points. Each sample selection iteration consisted of selection of 20 samples and the data were modeled with Kriging. The total budget of samples allotted to each run was 1000 samples for the Ackley function, 850 for the bird function and 750 for the Branin function. A total of four independent runs were performed for each scheme and the average of the true model error (RRSE) is compared to the LOLA-Voronoi algorithm in Table 1.

Note that the value of ε for the ε -greedy approach was taken to be 0.7, 0.8 and 0.9 respectively for the Ackley, Bird and Branin functions. For the ε -decreasing approach, the initial value of ε was set to 1 and it was decreased by 2% in each iteration. The choice of the value depends upon the non-linearity of the problem, and is largely based on the practitioner's understanding of the problem.

Table 1: Comparison of model error for LOLA-Voronoi, ε -greedy, ε -decreasing and ε -adaptive approaches.

Function	LOLA-Voronoi	ε -greedy	ε -decreasing	ε -adaptive
Ackley Function	0.09797	0.08785	0.08613	0.02980
Bird Function	1.23605	1.08317	0.66860	0.63585
Branin Function	1.907×10^{-7}	1.861×10^{-7}	1.986×10^{-7}	2.121×10^{-7}

The ε -greedy and ε -decreasing approaches perform slightly better than LOLA-Voronoi, hereby reaffirming the belief that controlling the extent of exploration and exploitation can lead to a more accurate model. It is however found that the results of these schemes is highly dependent on the value of ε , which is usually not known up-front. This makes the performance rather sensitive and prone to manual tuning.

It was found that the ε -adaptive approach achieves error values that are smaller than the other schemes, albeit at the expense of building intermediate models. For the Ackley function, it was observed that till approximately 800 samples, the error estimates of both schemes were similar and thereafter the exploitation component of the ε -adaptive approach got increasing portions of new samples. Thus, with only 200 additional samples, ε -adaptive approach managed to halve the error rate. The results are in line with the expected behavior of starting with more exploration and ending with more exploitation as the error rate goes down. It can also be seen from Fig. 2 that the algorithm samples more densely at the slopes than LOLA-Voronoi. The results for the Bird function, which also exhibits non-linear behavior are similar to the Ackley function, and the ε -adaptive schemes outperforms LOLA-Voronoi by substantial margins. The results for Branin function are comparable since the function is relatively smooth and therefore less sensitive to the actual trade-off between both criteria.

5 CONCLUSIONS

In this paper, a sequential design strategy is considered where different balancing schemes are investigated to find a good trade-off between exploration of the design space and exploitation of dynamic regions. Deciding the amount of exploration and exploitation is not a trivial task, as it strongly depends on the

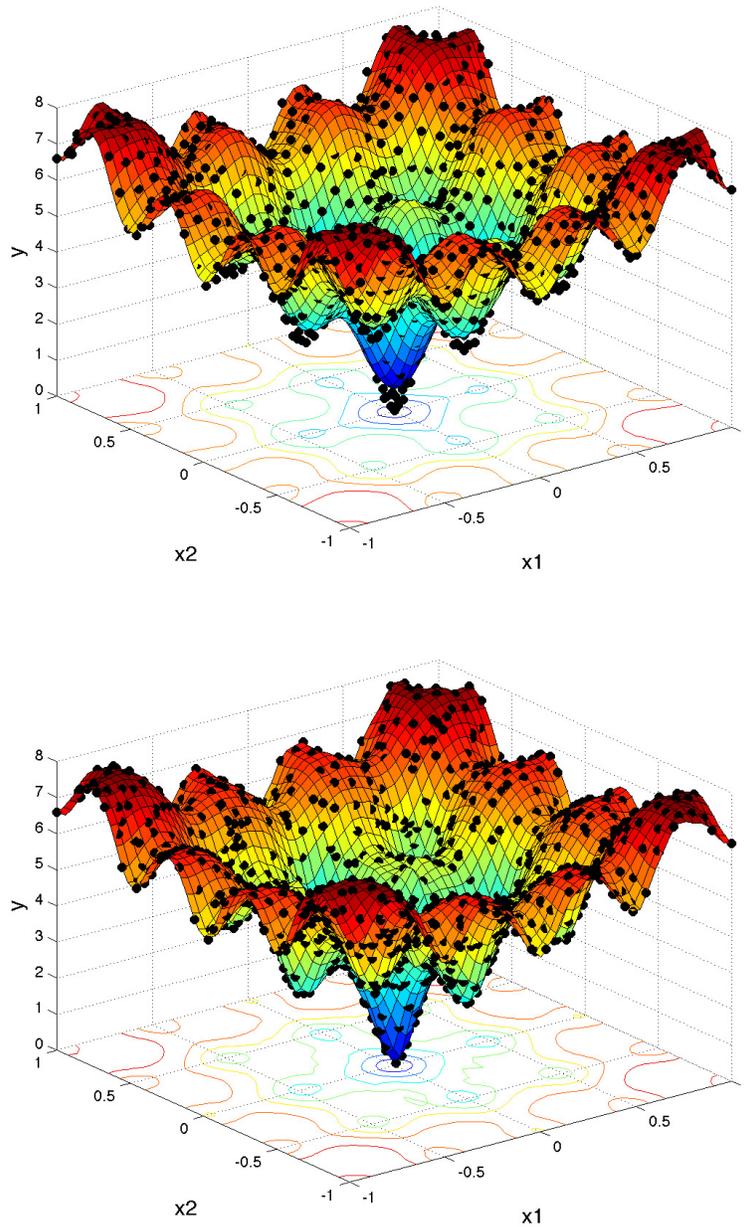


Figure 2: The top figure shows samples (dots) as selected by LOLA-Voronoi. The bottom figure shows samples (dots) as selected by ϵ -adaptive, where more samples can be seen at the slopes.

character of the function/problem being modeled. It is shown that adaptive balancing strategies can lead to models with a smaller error at the expense of building surrogate models during the sampling process.

ACKNOWLEDGMENTS

This research has been funded by the Interuniversity Attraction Poles Programme BESTCOM initiated by the Belgian Science Policy Office, and the Research Foundation Flanders (FWO-Vlaanderen). Dirk Deschrijver is a post-doctoral researcher of FWO-Vlaanderen.

REFERENCES

- Aerts, S., D. Deschrijver, W. Joseph, L. Verloock, F. Goeminne, L. Martens, and T. Dhaene. 2013. "Exposure assessment of mobile phone base station radiation in an outdoor environment using sequential surrogate modeling". *Bioelectromagnetics* 34 (4): 300–311.
- Crombecq, K., D. Gorissen, D. Deschrijver, and T. Dhaene. 2011. "A Novel Hybrid Sequential Design Strategy for Global Surrogate Modeling of Computer Experiments". *SIAM J. Scientific Computing* 33 (4): 1948–1974.
- Crombecq, K., E. Laermans, and T. Dhaene. 2011. "Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling". *European Journal of Operational Research* 214 (3): 683 – 696.
- Deschrijver, D., K. Crombecq, H. M. Nguyen, and T. Dhaene. 2011. "Adaptive sampling algorithm for macromodeling of parameterized S-parameter responses". *IEEE Transactions on Microwave Theory and Techniques* 59 (1): 39–45.
- Deschrijver, D., F. Vanhee, D. Pissoort, and T. Dhaene. 2012. "Automated Near-Field Scanning Algorithm for the EMC Analysis of Electronic Devices". *Electromagnetic Compatibility, IEEE Transactions on* 54 (3): 502–510.
- Gorissen, D., I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq. 2010, August. "A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design". *J. Mach. Learn. Res.* 99:2051–2055.
- Lehmensiek, R., and M. M. Petrie Meyer. 2002. "Adaptive sampling applied to multivariate, multiple output rational interpolation models with application to microwave circuits". *International Journal of Rf and Microwave Computer-aided Engineering* 12:332–340.
- Wang, G. G., and S. Shan. 2007. "Review of Metamodeling Techniques in Support of Engineering Design Optimization". *Journal of Mechanical Design* 129 (4): 370–42.

AUTHOR BIOGRAPHIES

PRASHANT SINGH received his B.Sc. (H) and M.Sc. degrees in Computer Science from the University of Delhi, India in 2009 and 2011 respectively. From July 2011 till September 2012, he worked as a Software Engineer at Nagarro Software. Since October 2012, he is active as a Ph.D. student in the research group Internet Based Communication Networks and Services (IBCN) at Ghent University. His research interests include surrogate modeling, machine learning, scientific computing and high performance computing. His email address is prashant.singh@ugent.be.

DIRK DESCHRIJVER was born in Tielt, Belgium, on September 26, 1981. He received the Masters degree (licentiaat) and the Ph.D. degree in computer science from the University of Antwerp, Antwerp, Belgium, in 2003 and 2007, respectively. He was a Marie Curie Fellow with the Scientific Computing Group, Eindhoven University of Technology, Eindhoven, The Netherlands, from May to October 2005. He is currently an FWO PostDoctoral Research Fellow with the Department of Information Technology, Ghent University, Ghent, Belgium. His current research interests include robust parametric macromodeling, rational least-squares approximation, orthonormal rational function, system identification, and broadband macromodeling techniques. His email address is dirk.deschrijver@ugent.be.

TOM DHAENE received the Ph.D. degree in electrotechnical engineering from Ghent University, Ghent, Belgium, in 1993. He was a Research Assistant with the Department of Information Technology, Ghent

University, from 1989 to 1993, where his research focused on different aspects of full-wave electromagnetic circuit modeling, transient simulation, and timedomain characterization of high-frequency and highspeed interconnections. He joined the EDA Company Alphabit (now part of Agilent) in 1993. He was one of the key developers of the planar EM simulator ADS Momentum, ADS Model Composer, and ADS Broadband SPICE. Since 2007, he has been a Full Professor with the Department of Information Technology, Ghent University - Institute for Broadband Technology. He has authored or coauthored more than 250 peer-reviewed papers and abstracts in international conference proceedings, journals, and books. He holds five U.S. patents. His email address is tom.dhaene@ugent.be.