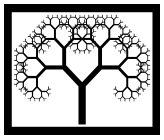
Paper 52



Proceedings of the First International Conference on
Soft Computing Technology in
Civil, Structural and Environmental Engineering,
B.H.V. Topping and Y. Tsompanakis, (Editors),
Civil-Comp Press, Stirlingshire, Scotland

Automated Response Surface Model Generation with Sequential Design

I. Couckuyt[1] K. Crombecq[2], D. Gorissen[1] and T. Dhaene[1]
[1] Department of Information Technology (INTEC-IBBT)
Ghent University, Belgium
[2] Department of Computer Science
University of Antwerp, Belgium

Abstract

The increasing use of expensive computer simulations in engineering places a serious computational burden on associated optimization problems. Surrogate modelling becomes standard practice in analyzing such expensive blackbox problems. Moreover, surrogate based optimization (SBO) is able to drastically reduce the number of needed function evaluations with respect to traditional methods. This paper briefly discusses several approaches available which use surrogate models for optimization and highlights one sequential design approach in particular, i.e., expected improvement. Expected improvement is described in detail, along with recent related work. The approach has been implemented in a readily available research platform for surrogate modelling and demonstrated on a concrete application from Electro-Magnetics (EM). The results hold competitive designs and one optimum is even able to outperform the reference optimum obtained using extensive domain specific knowledge.

Keywords: expected improvement, surrogate model, metamodel, optimization, sequential design, adaptive sampling, application, electro-magnetics.

1 Introduction

For many problems from science, and engineering it is impractical to perform experiments on the physical world directly (e.g., airfoil design, earthquake propagation). Instead, complex, physics-based simulation codes are used to run experiments on computer hardware. While allowing scientists more flexibility to study phenomena under controlled conditions, computer experiments require a substantial investment of computation time. This is especially evident for routine tasks such as optimization, sensitivity analysis and design space exploration [41].

As a result researchers have turned to various approximation methods that mimic the behavior of the simulation model as closely as possible while being computationally cheap(er) to evaluate. This work concentrates on the use of data-driven approximations using compact surrogate models (also known as metamodels, or response surface models (RSM)). Examples of surrogate models include: rational functions, Gaussian Process (GP) models, and Support Vector Machines (SVM).

It is crucial to stress the distinction between local and global surrogate modelling. With the latter the surrogate model itself is the goal, i.e., constructing a global surrogate model that attempts to imitate the response of the real simulation model as closely as possible over the whole design space. The resulting global surrogate model is able to fully replace the expensive simulator in design and analysis. In contrast, local surrogate models used in optimization only seeks one or more optima, the surrogate model itself is often discarded as it contains little extra information.

Constructing accurate surrogate models as efficiently as possible is an entire research domain in itself. In order to come to an acceptable model, numerous problems and design choices need to be overcome (what data collection strategy to use, what model type is most applicable, how should the model parameters be tuned, which variables are relevant, how to integrate domain knowledge, etc.). This work describes a popular optimization method for expensive blackbox simulators based on surrogate modelling, i.e., expected improvement. This approach has been implemented as a data collection (=sequential design) strategy in a flexible research platform for surrogate modelling, the **SU**rrogate **MO**delling (SUMO) Toolbox [13]. Consequently allowing different model types and model selection methods to be used without additional effort. Hence, the expected improvement implementation has been tested with 3 different modelling strategies on an optimization problem from Electro-Magnetics (EM).

Section 2 highlights critical components in surrogate modelling. In the next section the expected improvement approach and related work is presented, i.e., section 3. This approach is demonstrated to design an inter-digital filter. Details of the application are found in section 4, while the experimental setup is described in section 5. Results and conclusion form the last 2 sections of this paper, i.e., sections 6 and 7.

2 Surrogate modelling

As stated in the introduction, the principal reason driving surrogate model use is that the simulator is too time consuming to run for a large number of simulations. One model evaluation may take many minutes, hours, days or even weeks [41]. A simpler approximation of the simulator is needed to make optimization, design space exploration, etc. feasible.

An important aspect of surrogate modelling is data collection. Since data is computationally expensive to obtain, it is impossible to use traditional, one-shot, space filling designs. Data points must be selected iteratively, there where the information

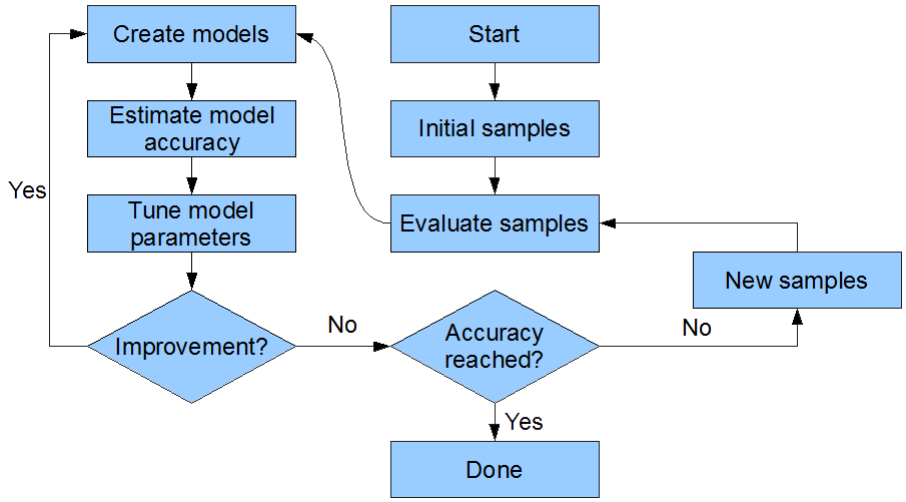


Figure 1: Flow chart of the SUMO toolbox

gain will be the greatest. A sampling function is needed that minimizes the number of sample points selected in each iteration, yet maximizes the information gain of each iteration step. This process is called sequential design, but is also known as active learning, or adaptive sampling. Other aspects of surrogate modelling include choosing the right type of approximation model for the problem at hand, a hyperparameter tuning strategy and a performance measure to asses the accuracy of the model.

The primary users of global surrogate modelling methods are application scientists, few of which will be experts in the intricacies of efficient sampling and modelling strategies. Their primary concern is obtaining an accurate replacement metamodel for their problem as fast as possible and with minimal overhead. Model (type) selection, model parameter optimization, sampling strategy, etc. are of lesser or no interest to them. At the same time every discipline or engineer has their preferred technique and approximation method and there is no such thing as a "one size fits all".

The authors argue a more holistic, integrated approach is needed to deal with the increasing diversity and computational complexity of current simulation codes. This has of course been noticed before (a discussion is given in [22]). Various authors have proposed powerful and novel sequential replacement metamodelling techniques, e.g., [14, 6, 11]. However, there is still room for a more flexible, and automated approach to surrogate modelling, that does not mandate assumptions (but does not preclude them either) about the problem, model type, sampling algorithm, etc. Particularly, if such assumptions are neither applicable nor practical.

The software used in this paper is the SUMO Toolbox which is an adaptive tool that integrates different modelling approaches and implements a fully automated, adaptive global surrogate model construction algorithm. The general flow of the toolbox is illustrated in figure 1. First, an experimental design, often from Design of Experiments (DOE), is specified and evaluated. Subsequently, surrogate models are built to fit this

data as good as possible, according to a set of measures (e.g., cross validation). The model parameter (hyperparameters) are estimated using an optimization algorithm. The accuracy of the set of surrogate models is improved until no further improvement can be obtained (or when another stopping criterion such as time is met). If all stopping criteria are satisfied the process is halted and the final, best surrogate model is returned. On the other hand, when at least one criterion is not met, a sequential design strategy will select new samples to be evaluated and the surrogate models are updated with this new information. Given a simulation engine the toolbox is able to produce an accurate global surrogate model with as little user interaction as possible.

While the focus of the toolbox is thus on global surrogate models the toolbox is easily adapted to perform optimization of the expensive simulator. In fact, the optimization method discussed in section 3 is a data collection strategy and thus can simply be implemented as a new sequential design method. Other components of the toolbox (approximation methods, hyperparameter tuning strategies, performance measures, etc.) can be reused.

3 Expected improvement

Surrogate based optimization (SBO) is an important research domain to accelerate optimization of expensive simulation problems. Existing optimization algorithms have been adapted to reduce the number of function evaluations as well as utilizing parallel computing to speedup the optimization process. A good overview is given by [2], however, optimization methods can still greatly benefit by taking advantage of surrogate models. The extra information gained helps in avoiding local optima and efficiently guiding the search to global optima. E.g., a surrogate model offers a cheap alternative to the expensive simulator for exploration and exploitation purposes, likewise the practitioner is able to explore the region of the final optimum quite easily, etc. Various directions have been taken to incorporate surrogate models in the optimization process.

Trust region frameworks manage a local surrogate model throughout the design space. The move limit strategy, i.e., a set of mathematical based or pure heuristic rules, determines the size and location of the surrogate model. For more information the reader is referred to Alexandrov et al. [1]. While trust region frameworks are widely used in large scale optimization problems they have the disadvantage of often overlooking the global optimum, as only a very small part of the design space is approximated by the local surrogate model. On the other hand, by approximating only one local part of the design space at a time, it is possible to optimize very difficult, non-linear blackbox functions.

Another, more recent, approach is to use global surrogate models and emphasize more on adaptive sampling. However note that while these surrogate models are a global approximation, they are not necessarily accurate over the whole design space, which depends on the adaptive sampling strategy. In engineering, adaptive sampling

strategies are also known as infill criteria. Starting from an initial approximation of the design space, the infill criterion identifies new samples of interest (infill or update points) to update the approximation model. It is crucial in global SBO to strike a correct balance between exploration[1] and exploitation[2]. A well-known infill criterion that is able to effectively solve this trade-off is expected improvement, popularized by Jones et al. [17, 35] in the Efficient Global Optimization (EGO) algorithm. Although, to the best of the author's knowledge expected improvement has been suggested in literature as early as 1978 [27]. Jones wrote an excellent discussion to the infill criteria approach in [16]. Subsequently, Sasena compared different infill criteria for optimization and investigated extensions of those infill criteria for constrained optimization problems in [34].

The expected improvement equation (4) is more easily interpreted graphically, as in figure 2. At $\overrightarrow{x} = 0.5$ a Gaussian probability density function is shown that express uncertainty about the predicted function value. Thus, the uncertainty at any point $\overrightarrow{x}$ is treated as the realization of a random variable $Y(\overrightarrow{x})$ with mean $\hat{y} = \hat{f}(\vec{x})$ (=prediction) and variance $\hat{s}^2 = \hat{\sigma}(\vec{x})$ (=prediction variance). Assume the random variable is normally distributed, then the shaded area is the probability of improvement of any newly calculated function value $y(\overrightarrow{x})$ over the current minimum function value f_{min}(the dotted line), denoted as $P(y < f_{min})$ where,

$$\begin{aligned} P(y \leq f_{min}) &= \int_{-\infty}^{f_{min}} Y(\overrightarrow{x})\, dy \\ &= \Phi\left(\frac{f_{min} - \hat{y}}{\hat{s}}\right), \end{aligned} \tag{1}$$

$\Phi(\cdot)$ is the standard normal cumulative distribution function. The probability of improvement is already a very handy infill criterion. Although, this criterion describes the possibility of a better minimum function value, it does not quantify how large this improvement will be.

Expected improvement quantifies the improvement and is the first moment of the shaded area, i.e., every possible improvement over f_{min} multiplied by the associated likeliness. For continuous functions this is an integral defined by,

$$I = max(f_{min} - y, 0), \tag{2}$$

$$E(I) = \int_{-\infty}^{f_{min}} I \cdot Y(\overrightarrow{x})\, dy. \tag{3}$$

[1] enhancing the general accuracy of the surrogate model

[2] enhancing the accuracy of the surrogate model solely in the region of the (current) optimum

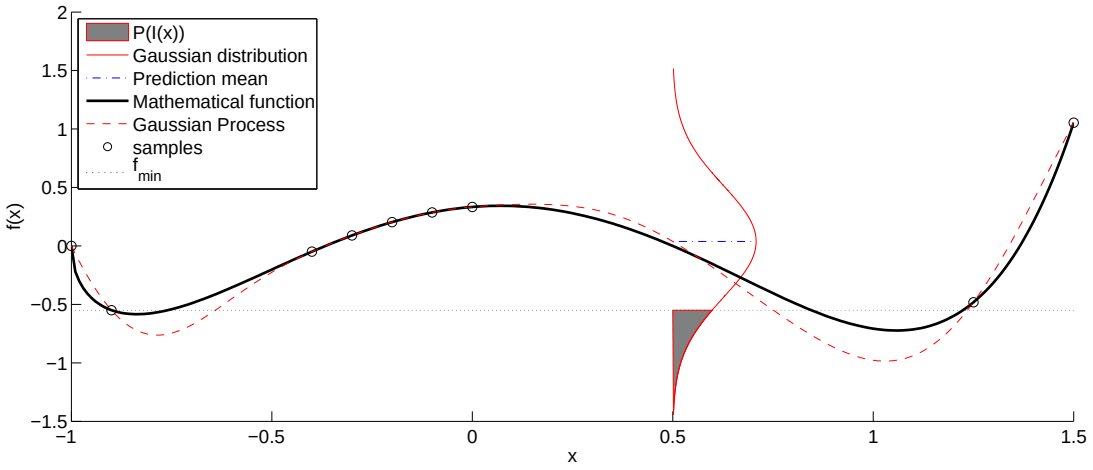


Figure 2: Graphical illustration of Gaussian Processes and E[I(x)]

This can be further developed and the following closed formula is obtained for expected improvement,

$$E(I) = \begin{cases} (f_{min} - \hat{y}) \cdot \Phi\left(\frac{f_{min}-\hat{y}}{\hat{s}}\right) + \hat{s} \cdot \phi\left(\frac{f_{min}-\hat{y}}{\hat{s}}\right) & if\ \hat{s} > 0 \\ 0 & if\ \hat{s} \leq 0 \end{cases}, \tag{4}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal probability density function and standard normal cumulative distribution function, respectively. Expected improvement and probability of improvement serve as utility functions, often conceived as figures of merit, which have to be optimized over $\overrightarrow{x}$ to find the subsequent sample point to evaluate. Note, however, that besides the prediction $\hat{y} = \hat{f}(\overrightarrow{x})$ of the surrogate model, a point-wise error estimation $\hat{s}^2 = \hat{\sigma}(\overrightarrow{x})$ is needed.

Therefore, the original EGO algorithm used kriging as surrogate model of choice, as kriging provides closed formulae for prediction as well as a point-wise error estimation. Unfortunately, a full mathematical description of kriging is out of scope for this paper. Kriging has been explained many times in the literature, hence, a small overview of the most influential papers is given here. A good starting point for kriging would be the introductions of Matheron et al. and Sacks et al. [26, 33]. Kriging, in fact, is part of a broader class of approximation methods, i.e., Gaussian Processes (GP). While traditional approximation methods only predict a single function value, GP methods predict a complete normal distribution $Y(\vec{x}) \sim \mathcal{N}(\hat{y}, \hat{s})$ for a point $\vec{x}$. The predicted distribution imparts the probability that a particular function value occurs. Generally, in Bayesian analysis, it is also possible to use a stochastic process as prediction model, allowing any kind of distribution function. However, solving such approximation models require expensive (and sophisticated) Markov Chain Monte Carlo simulation (MCMC) (for an implementation, see [25]) or a, more deterministic, variational technique [3].

A good insight into expected improvement using a simple 1D approximation model

(a so-called Wiener process) is given by Locatelli [23]. However, for a full overview of modern GP the reader is referred to Gibbs et al. [12], and, the more recent, GP reference book of Rasmussen et al. [29]. Depending on the context some authors coin the term Gaussian process (temporal) or Gaussian Random Field Metamodels (GRFM; spatial) [7], however, the underlying methods are the same. As a final note, in case there is a cheaper, but more inaccurate, simulator available multi-fidelity approximation models can be used. For instance, co-kriging is able to incorporate several levels of fidelity samples in its prediction [4, 19]. This has been successfully used to optimize a transonic civil aircraft wing in [9].

Over the years small changes and improvements to expected improvement have been suggested. Sobester et al. [38] added a weight w to the 2 terms comprising the expected improvement equation, intuitively referring to it as weighted expected improvement (see equation 5). The weights allow fine-tuning of the balance between exploitation and exploration. In addition, Radial Basis Functions (RBF) were used instead of kriging.

$$wE(I) = \begin{cases} w \cdot (f_{min} - \hat{y}) \cdot \Phi\left(\frac{f_{min}-\hat{y}}{\hat{s}}\right) + (1-w) \cdot \hat{s} \cdot \phi\left(\frac{f_{min}-\hat{y}}{\hat{s}}\right) & if\ \hat{s} > 0 \\ 0 & if\ \hat{s} \leq 0 \end{cases}. \quad (5)$$

Parallel versions of EGO have been proposed by [37, 28]. While expected improvement is a very popular criterion, other infill criteria were suggested that solve the exploitation and exploration trade-off completely different. For instance, Regis et al. introduced the Constrained Optimization using Response Surfaces (CORS) method in [30].

The prediction variance of kriging is known to underestimate the true error, this sometimes causes expected improvement to put too much stress on exploiting a single region of interest. Therefore, expected improvement has been extended to sample more conservatively, stressing the global exploration when the number of samples is low and tends to the normal expected improvement when the number of samples increase [10]. Another approach is to use (expensive) bootstrapping to estimate the prediction variance more accurately, as done by den Hertog et al. [5].

While infill criteria are often optimized to find the best infill points there are other, less expensive, search strategies available. Wang et al., for instance, treat the infill criterion as a probability density function and propose a dimensional-free method to efficiently sample according to the probability given by the infill criterion. This has been implemented by considering the prediction itself as the infill criterion in [42, 36], thus focusing only on exploitation, though other infill criteria may be used.

Another topic that has been the subject of extensive research is that of multi-objective surrogate based optimization (MOSBO). Expected improvement has been used in a multi-objective setup in various ways. ParEGO [20] uses a dynamic weighted sum of the objective functions. This aggregate function is approximated using kriging, afterwards expected improvement is used to determine the new point to be evaluated.

The dynamic part of the weighted sum involves randomly changing the weights every iteration. Thus, every iteration progress is made to identify a single optimum on the pareto front. After many iteration, this results in several optima which approximately cover the pareto front. Another approach is to apply the expected improvement concept directly to the Pareto front [18, 40]. Using this approach two kind of improvements can be distinguished. Points that augment the current front and points that improve on the front, which in turn consist of multiple levels of improvement (i.e., on how many points of the frontier it improves on). Other MOSBO techniques consist of hybrids between classic optimization techniques and surrogate models, e.g., the surrogate based variants of NSGA-II [40]. Although the research into MOSBO is still very young, an excellent overview of current research is already available in [21] and [8].

4 Application

The studied application in this paper is a common type of inter-digital filters broadly used in antennae of cellular phones, speech recognition, etc. There are various ways to analyze such electronic circuits. Circuit simulation allows for very fast evaluation with reasonable accuracy, while finite element methods (FEM) provide high accuracy but are also the most expensive. The inter-digital filter presented in this paper has been solved by a domain expert, Swanson [39], using a combination of analytic methods, circuit simulation and FEM. For the sake of illustration the finite element model presented herein is used to illustrate expected improvement, trying to achieve similar results as a domain expert.

The commercial CST MICROWAVE STUDIO® (CST MWS) software from Computer Simulation Technology (CST) is used for the full 3D simulation. CST MWS is a popular high frequency EM simulation tool widely used in the industry. To reproduce the finite element model in CST MWS a similar approach to the one of Swanson was used. A simplification of the 3D model is depicted as a top-down view in figure 3 and consists of 5 microstrip resonators layed out in parallel. The finite element model is fully parametrized by $S1 = [32, 38]$, $S2 = [40, 48]$, $off1 = [-3, 9]$, $off2 = [-3, 3]$ and $off3 = [-3, 3]$ (the numbers in the brackets are the bounds). The last 3 parameters define the offsets in millimeter of the strips with respect to the parallel part (dotted lines) of the neighboring strip. In particular, $off1$ is the offset of the outer 2 strips, $off2$ of the second and fourth strip and $off3$ of the middle strip. In other words, the offsets implicitly define the length of each microstrip. While $S1$ and $S2$ (spacings) denote the gap between the outer 2 strips and the inner 2 strips, respectively. In sum, this adds up to a total of 5 geometric design variables that need to be optimized. The structure is symmetric and is constructed so that, for any given set of parameters, the design is modified to accommodate the other components.

The goal is to design a narrow-band filter which maximizes throughput centered on 2.44 GHz, with a 10% band-width. For practical purposes this is equal to minimizing the maximum of the S_{11}-parameter curve, i.e., the reflection coefficient, in the

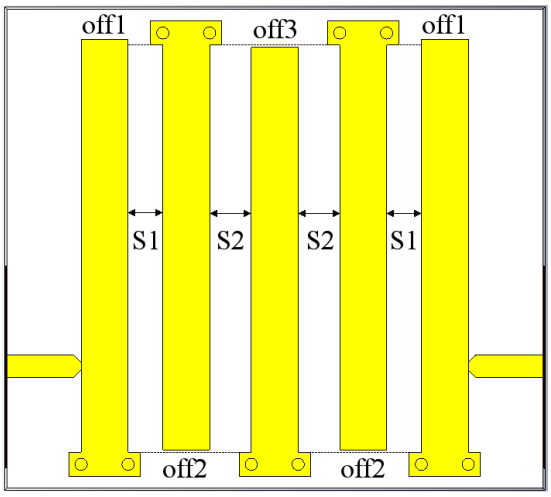


Figure 3: Layout view of the 3D narrow-band filter

frequency range $[2.32, 2.56]$ GHz. The finite element model is solved using the 'fast S-parameter' solver in CST MWS and requires approximately 5 to 10 minutes.

5 Experimental setup

Version 6.1 of the SUMO toolbox is utilized to perform the optimization of the narrow-band filter and is configured as followed. The initial set of samples is generated by a maximin Latin Hypercube Design (LHD; implemented as in [31]) of 19 points together with 32 corner points, adding up to a total of 51 initial points. Several variants of expected improvement are available in the toolbox, but for this particular application the original expected improvement function as defined in section 3 is used to select infill points. The expected improvement function is optimized using the DIviding RECTangles (DIRECT) algorithm of Jones et al. [15] to determine the next sample point to evaluate. A time budget constraint of 24 hours is applied, i.e., the total process runs for 24 hours.

The aforementioned configuration is reproduced 3 times with different surrogate modelling strategies. The first 2 cases configure kriging as surrogate model of choice as implemented by the DACE toolbox [24]. More precisely, one time the DACE toolbox itself performs the hyperparameter optimization, which comprises maximum likelihood estimation (MLE) using pattern search. In the other run, the hyperparameters of the kriging model are identified by Matlab's Genetic Algorithm (GA) toolbox using 10-fold cross validation to guide the search. In the last, and final configuration, a custom implementation of blind kriging [32] is employed (MLE using the DIRECT algorithm). Optimization of the hyperparameters θ occur in log space with lower bounds $\theta_{lower} = (-5,-5,-5,-5,-5)$, upper bounds $\theta_{upper} = (3,3,3,3,3)$ and starting point $\theta_0 = (0,0,0,0,0)$. The kriging models based on the DACE toolbox use a linear regression function and all kriging models utilize the standard Gaussian corre-

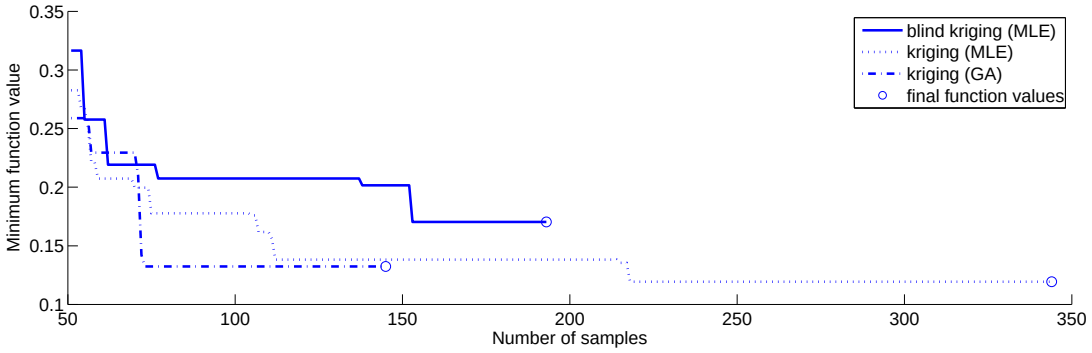


Figure 4: Evolution of the minimum function value versus the number of samples evaluated

lation function.

6 Results

Figure 4 shows the progress of the optimization process, i.e., current minimum function value versus the number of samples evaluated. The solid line is the blind kriging configuration and performs worst in terms of the final optimum. This may look counter-intuitive, as blind kriging is a more powerful approximation model, but is easily explained by the fact that fitting a blind kriging model is twice as expensive as standard kriging, hence, less time is available to evaluate the expensive simulation code. For this particular problem the benefit of investing in evaluating extra samples is greater than the approximation method itself. On the other hand, kriging (GA) is able to produce better kriging models due to a large search with a genetic algorithm guided by cross validation, and, thus, that configuration finds attracting basins more quickly. However due to the cost of cross validation and evolutionary based strategies it is only able to evaluate approximately 190 samples before the time budget is exceeded, however, at that point it is the best performing method. The standard kriging configuration finds the best optimum. As it is significantly faster than the other 2 configurations, it is able to process more simulator runs, which, as said above, seems more important in this application than a really accurate approximation model. After about 220 function evaluations it finds the final optimum and still has time to select about 100 more samples to look for an even better optimum or validate the current one. It should be noted that the usefulness of more expensive surrogate modelling strategies (such as blind kriging) may improve when the time to run a simulation increases to hours or even days.

For reference purposes the optimal design found by Swanson [39] (see section 4) is included in the comparison. While it is hard to outperform methods that use domain specific knowledge it can still serve as a baseline to compare the results. The final optima found using expected improvement are displayed in table 1, together with the reference optimum.

method	$\|X\|$	$\vec{x_{min}}$	f_{min}
Blind kriging (MLE)	193	$(35.03, 41.37, 8.99, -1.14, 0.19)$	0.17038
Kriging (GA)	145	$(36.00, 43.10, 5.90, -3.00, -3.00)$	0.13234
Kriging (MLE)	344	$(36.68, 44.16, 6.16, -2.67, -2.47)$	0.11936
reference	*unknown*	$(37.10, 44.46, 6.30, -2.60, -2.43)$	0.12527

Table 1: Final designs of the inter-digital filter. $|X|$ is the number of samples evaluated, $\vec{x_{min}}$ and f_{min} are the final optima and function values respectively

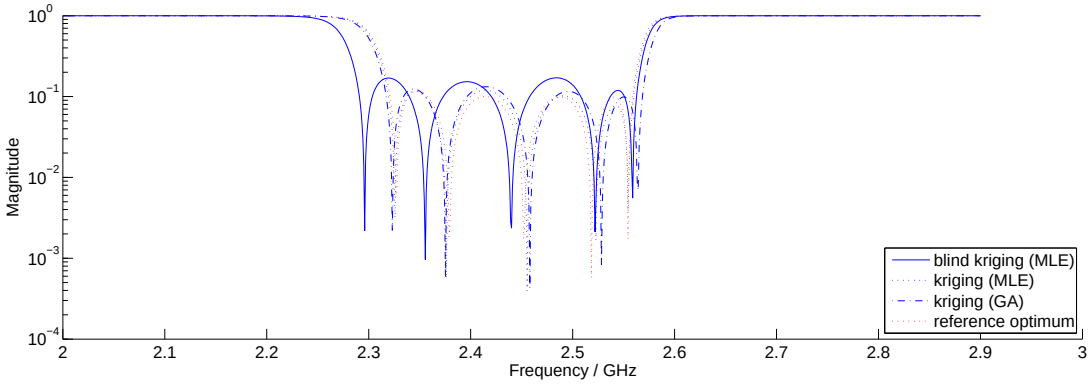


Figure 5: S_{11}-parameter magnitude plots

Surprisingly, the optimal design found by the kriging (MLE) configuration is found to perform better than the reference design, i.e., with respect to the same objective function. To compare the different designs fairly the S_{11}-parameter curves are drawn in figure 5. The reference design is constructed so that the ripples in the S_{11}-parameter curve are of equal height (exact equal ripple tuning). This is important so the filter has consistent performance over the whole frequency range of interest. However, regard that the actual goal function used is the maximum of the S_{11}-parameter curve in the frequency range $[2.32, 2.56]$ GHz and equal ripple tuning has not been factored in. Thus, the optimum found in this paper is better on the goal function but has not exact equal ripples (though very close to), while the reference design is slightly more consistent on the whole frequency with regard to the ripples.

A large advantage of using surrogate models to aid optimization is the ability to easily explore the final (and intermediate) approximation models. The practitioner is able to cheaply analyze the robustness of the optimum, locate other interesting regions (e.g., local optima), etc. For illustration purposes the region of the best optimum of the kriging (MLE) run has been located and a plot is shown in figure 6. This plot is a 2 dimensional slice of the full design space where the offset parameters are set constant according to the best optimum, i.e., $off1 = 6.16$, $off2 = -2.67$ and $off3 = -2.47$. As expected this region is densely sampled as the figure of merit, i.e., expected improvement, exhaustively explored this promising region. The samples are denoted by the dots, the larger the dot the closer it is to the actual slice.

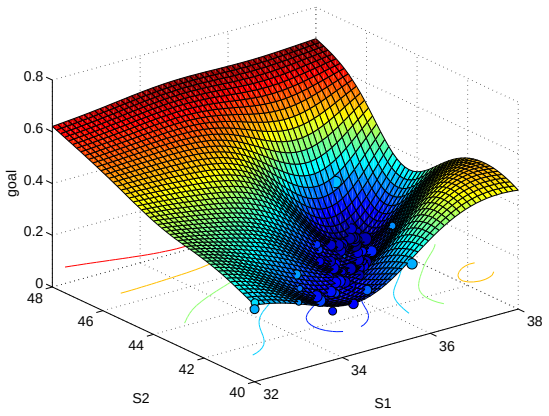


Figure 6: Region around the final design of Kriging (MLE)

7 Conclusion and future work

The main contribution of this work is the adaptive integration of different state-of-the-art surrogate modelling components (sample selection, model tuning, ...) into a fully functional, flexible software implementation, applicable to a very wide range of domains, from global surrogate modelling to optimization of expensive simulation codes. This work illustrates this by integrating a popular optimization approach into a research platform, the SUMO toolbox, which focuses on global surrogate modelling. The result is a powerful set of methods which can be used whether the practitioner needs a global accurate approximation or optimization of the simulation code.

This optimization approach, i.e., expected improvement, has been thoroughly explained together with related work. Subsequently, expected improvement is used to optimize a narrow-band filter problem with no domain specific knowledge. The results are in comparison to a reference design obtained by an expert in the field and one test configuration is even able to improve upon this reference design.

Future work includes solving more real-world examples (e.g., antenna problems) and researching multi-objective surrogate based optimization (MOSBO) methods.

Acknowledgments

Ivo Couckuyt is funded by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The authors like to thank Computer Simulation Technology AG (CST AG; http://www.cst.com) and in particular Frank Mosler for their help with the application and the CST MICROWAVE STUDIO® software.

References

[1] N. Alexandrov, J.E. Dennis, R.M. Lewis, V. Torczon, "A Trust Region Framework for Managing the Use of Approximation Models in Optimization", Technical Report NASA CR-201745 ICASE Report No. 97-50, Institute for Computer Applications in Science and Engineering, October 1997.

[2] S. Andradóttir, "A Review of Simulation Optimization Techniques", in *Proceedings of the Winter Simulation Conference*, pages 151–158. IEEE Computer Society Press, 1998.

[3] C.M. Bishop, M.E. Tipping, *Bayesian Regression and Classification*, Volume 190, Chapter 13, pages 267–288, IOS Press, 2003.

[4] N.A.C. Cressie, *Statistics for spatial data*, John Wiley & Sons, 1993.

[5] D. den Hertog, J. Kleijnen, A. Siem, "The correct Kriging variance estimated by bootstrapping", *Operational Research Society*, 57: 400–409, 2006.

[6] V. Devabhaktuni, B. Chattaraj, M. Yagoub, Q.J. Zhang, "Advanced microwave modeling framework exploiting automatic model generation, knowledge neural networks, and space mapping", *IEEE Tran. on Microwave Theory and Techniques*, 51(7): 1822–1833, 2003.

[7] M.T.M. Emmerich, K. Giannakoglou, B. Naujoks, "Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels.", *IEEE Trans. Evolutionary Computation*, 10(4): 421–439, 2006.

[8] A. Forrester, A. Sobester, A. Keane, *Engineering Design Via Surrogate Modelling: A Practical Guide*, Wiley, 2008, page 228.

[9] A.I. Forrester, A. Sobester, A.J. Keane, "Multi-fidelity optimization via surrogate modelling", *Royal Society*, 463(2088): 3251–3269, 2007.

[10] A.I.J. Forrester, D.R. Jones, "Global Optimization of Deceptive Functions With Sparse Sampling", in *Proceedings of the AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*. AIAA, 2008.

[11] M. Ganser, K. Grossenbacher, M. Schutz, L. Willmes, T. Back, "Simulation meta-models in the early phases of the product development process", in *Proceedings of Efficient Methods for Robust Design and Optimization (EUROMECH 07)*, 2007.

[12] M. Gibbs, D.J.C. Mackay, "Efficient implementation of Gaussian processes", Technical report, Department of Physics, Cavendish Laboratory, Cambridge University, 1997.

[13] D. Gorissen, L. De Tommasi, K. Crombecq, T. Dhaene, "Sequential Modeling of a Low Noise Amplifier with Neural Networks and Active Learning", *Neural Computing and Applications, accepted*, 2008.

[14] R.B. Gramacy, H.K.H. Lee, W.G. Macready, "Parameter space exploration with Gaussian process trees", in *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, page 45. ACM Press, New York, NY, USA, 2004, ISBN 1-58113-828-5.

[15] D. Jones, C. Perttunen, B. Stuckman, "Lipschitzian optimization without the Lipschitz constant", *Optimization Theory and Applications*, 79(1): 157–181,

1993.

[16] D.R. Jones, "A Taxonomy of Global Optimization Methods Based on Response Surfaces", *Global Optimization*, 21: 345–383, 2001.

[17] D.R. Jones, M. Schonlau, W.J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions", *J. of Global Optimization*, 13(4): 455–492, 1998, ISSN 0925-5001.

[18] A.J. Keane, "Statistical Improvement Criteria for Use in Multiobjective Design Optimization", *AIAA Journal*, 44(4): 879–891, 2006.

[19] M.C. Kennedy, A. O'Hagan, "Predicting the output from complex computer code when fast approximations are available", *Biometrika*, 87: 1–13, 2000.

[20] J. Knowles, "ParEGO: A Hybrid Algorithm With On-Line Landscape Approximation for Expensive Multiobjective Optimization Problems", *IEEE Transactions on Evolutionary Computation*, 10(1): 50–66, 2006.

[21] J. Knowles, H. Nakayama, "Meta-modeling in Multiobjective Optimization", in *Multiobjective Optimization - Interactive and Evolutionary Approaches*. Springer LNCS, In press, 2008.

[22] Y. Lin, *An Efficient Robust Concept Exploration Method and Sequential Exploratory Experimental Design*, PhD thesis, Georgia Institute of Technology, 2004.

[23] M. Locatelli, "Bayesian Algorithms for One-Dimensional Global Optimization", *Global Optimization*, 10: 57–76, 1997.

[24] S.N. Lophaven, H.B. Nielsen, J. Søndergaard, "Aspects of the Matlab toolbox DACE", Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2002.

[25] D.J. Lunn, A. Thomas, N. Best, D. Spiegelhalter, "WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility", *Statistics and Computing*, 10: 325–337, 2000.

[26] G. Matheron, "Principles of Geostatistics", *Economic Geology*, 58: 1246–1266, 1963.

[27] J. Mockus, V. Tiesis, A. Zilinskas, "The application of Bayesian methods for seeking the extremum", *Towards Global Optimization*, 2: 117–129, 1978.

[28] W. Ponweiser, T. Wagner, M. Vincze, "Clustered Multiple Generalized Expected Improvement: A Novel Infill Sampling Criterion for Surrogate Models", in *Congress on Evolutionary Computation*, 2008.

[29] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[30] R. Regis, C. Shoemaker, "Constrained Global Optimization of Expensive Black Box Functions Using Radial Basis Functions", *Journal of Global Optimization*, 31(1): 153–171, January 2005, ISSN 0925-5001, URL `http://dx.doi.org/10.1007/s10898-004-0570-0`.

[31] J.V. Roshan, H. Ying, "Orthogonal-Maximin Latin Hypercube Designs", *Statistica Sinica*, 18: 171–186, 2008.

[32] J.V. Roshan, H. Ying, S. Agus, "Blind Kriging: A New Method for Developing

Metamodels", *ASME Journal of Mechanical Design*, 130(3), 2008.

[33] J. Sacks, W.J. Welch, T. Mitchell, H.P. Wynn, "Design and analysis of computer experiments", *Statistical science*, 4(4): 409–435, 1989.

[34] M. Sasena, *Flexibility and Efficiency Enhancements For Constrainted Global Design Optimization with Kriging Approximations*, PhD thesis, University of Michigan, 2002.

[35] M. Schonlau, *Computer Experiments and Global Optimization*, PhD thesis, University of Waterloo, 1997.

[36] S. Shan, G.G. Wang, "An Efficient Pareto Set Identification approach for Multi-objective Optimization on Black-box Functions", *Mechanical Design*, 127(5): 866–874, 2005.

[37] A. Sóbester, S.J. Leary, A.J. Keane, "A parallel updating scheme for approximating and optimizing high fidelity computer simulations", *Structural and Multidisciplinary Optimization*, 27: 371–383(13), 2004.

[38] A. Sóbester, S.J. Leary, A.J. Keane, "On the Design of Optimization Strategies Based on Global Response Surface Approximation Models", *Global Optimization*, 33(1): 31–59, 2005, ISSN 0925-5001.

[39] D.G. Swanson, "Microwave Filter Design", *IEEE Microwave magazine*, 8: 105–114, 2007.

[40] I. Voutchkov, A. Keane, "Multiobjective Optimization using Surrogates", in I. Parmee (Editor), *Adaptive Computing in Design and Manufacture 2006. Proceedings of the Seventh International Conference*, pages 167–175. Bristol, UK, April 2006.

[41] G.G. Wang, S. Shan, "Review of Metamodeling Techniques in Support of Engineering Design Optimization", *Journal of Mechanical Design*, 129(4): 370–380, 2007.

[42] L. Wang, S. Shan, G.G. Wang, "Mode-Pursuing Sampling Method for Global Optimization on Expensive Black-box Functions", *Engineering Optimization*, 36(4): 419–438, 2004.