

Automatic Surrogate Model Building for Computer based Design

* Dirk Gorissen¹, Tom Dhaene¹, and Piet Demeester¹

¹ Ghent University
Gaston Crommenlaan 8 bus
201, 9050, Ghent, Belgium
dirk.gorissen@ugent.be

Key Words: *surrogate modeling, simulation, grid computing, applications*

ABSTRACT

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 (one simulation may take many minutes, hours, days or even weeks). This is especially evident for routine tasks such as optimization, sensitivity analysis and design space exploration [1].

As a result, the use of various approximation methods that mimic the behavior of the simulation model as closely as possible (while being computationally cheaper to evaluate), has become standard practice. This work concentrates on the use of data-driven, **global**¹ approximations using compact surrogate models (also known as metamodels, or response surface models (RSM)). Examples include: rational functions, Kriging models, and Support Vector Machines (SVM). Once they are constructed, global surrogate models provide a fast and efficient way for the engineer to explore the relationship between parameters (design space exploration), study the influence of various boundary conditions on different optimization runs, or enable the simulation of large scale systems where this would normally be too cumbersome. For the last case a classic example is the full-wave simulation of an electronic circuit board. Electro-magnetic modeling of the whole board in one run is almost intractable. Instead the board is modeled as a collection of small, compact, accurate replacement surrogate models that represent the different functional components (capacitors, resistors, ...) on the board. In this way simulations can be literally pieced together.

However, in order to come to an acceptable approximation, numerous problems and design choices need to be overcome: what data collection strategy to use, what model type is most applicable, how should model parameters be tuned, how to optimize the accuracy vs computational cost trade-off, etc. Particularly important is the data collection strategy. Since data is computationally expensive to obtain, data points must be selected iteratively, there where the information gain will be the greatest. A

¹Note the difference between global surrogate modeling as opposed to local surrogate modeling. In the global case, optimization is not the goal but rather a consequence. The accuracy requirements are also higher and the data collection strategy is different. Nevertheless the two are not disjunct, advances in one type can provide insights for the other.

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 adaptive sampling, but is also known as active learning, Optimal Experimental Design, and sequential design. Together this makes that there are an overwhelming number of options available to the designer: different model types, different experimental designs, different model selection criteria, etc.

However, in practice it turns out that the designer rarely tries out more than one subset of options. All too often, surrogate model construction is done in a one-shot manner. Iterative and adaptive methods, on the other hand, have the potential of producing a much more accurate surrogate at a considerably lower cost (less data points) [2]. We present a state-of-the-art research platform that provides an automatic, flexible and rigorous means to tackle such problems and that can easily be integrated in the engineering design process. The platform in question is the the **SURrogate MODELing Toolbox (SUMO)**.

The SUMO Toolbox is an adaptive tool that integrates different modeling approaches and implements a fully automated, adaptive global surrogate model construction algorithm. Given a simulation engine, the toolbox automatically generates a surrogate model within the predefined accuracy and time limits set by the user (see figure 1). However, at the same time keeping in mind that there is no such thing as a 'one-size-fits-all', different problems need to be modeled differently. Therefore the toolbox was designed to be modular and extensible but not be too cumbersome to use or configure. Different plugins

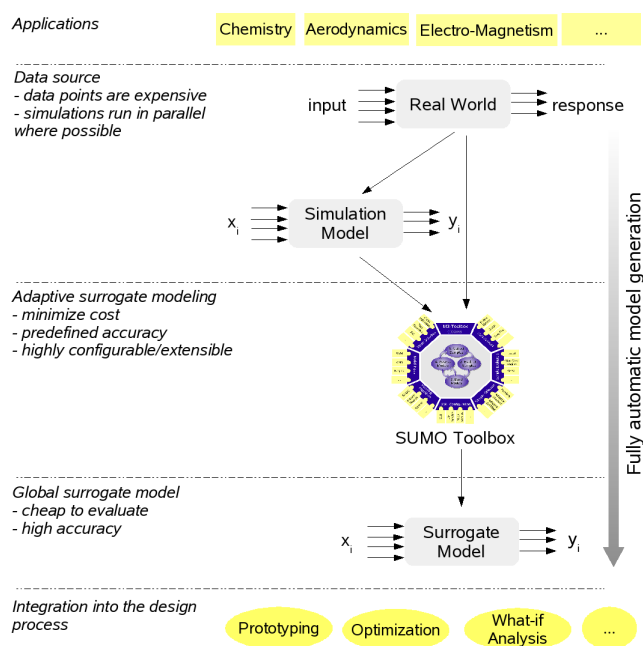


Figure 1: Automatic Adaptive Surrogate Modeling

are supported: model types (neural networks, Kriging, splines, ...), model parameter optimization algorithms (BFGS, GA, PSO, ...), adaptive sample selection (density based, gradient based, ...), and sample evaluation methods (local, on a cluster or grid). The behavior of each component is configurable through a central XML configuration file and components can easily be added, removed or replaced by custom, problem-specific, implementations.

Our approach has been successfully applied to a very wide range of fields ranging from combustion modeling in chemistry and metallurgy, semi-conductor modeling (Electro Magnetism), aerodynamic modeling (aerospace), to structural mechanics modeling in the Car industry. Its success primarily due to its flexibility, self tuning implementation, and its ease of integration into the larger computational science and engineering pipeline.

REFERENCES

- [1] G. Gary Wang and S. Shan. "Review of metamodeling techniques in support of engineering design optimization". *Journal of Mechanical Design*, Vol. **129**, 370–380, 2007.
- [2] Daniel Busby, Chris L. Farmer, and Armin Iske. "Hierarchical nonlinear approximation for experimental design and statistical data fitting". *SIAM Journal on Scientific Computing*, Vol. **29**, 49–69, 2007.