Meta-modeling with modeFRONTIER: Advantages and Perspectives

The progresses in finite elements methods (FEM) and high performance computing offer to engineers accurate and reliable virtual environments to explore various possible configurations. On the other hand and at the same time, the number of users' requests constantly increases going even beyond computational exhaustiveness.

In real case applications, it is not always possible to reduce the complexity of the problem and to obtain a model that can be solved quickly. Usually, every single simulation can take hours or even days. In such cases, the time frame required to run a single analysis, prohibits running more than a few simulations, hence other, smarter approaches are needed.

Engineers may consider and apply a Design of Experiment (DOE) technique to perform a reduced number of calculations. These well-distributed results can be subsequently used by the engineers to create a surface which interpolates these points. This surface represents a meta-model of the original problem and can be used to perform the optimization without computing any further analyses.

The use of mathematical and statistical tools to approximate, analyze and simulate complex real world systems is widely applied in many scientific domains. These types of interpolation and regression methodologies are now becoming common even in engineering where they are also known as Response Surface Methods (RSMs). RSMs are indeed becoming very popular as they offer a surrogated model with a second generation of improvements in speed and accuracy in computer aided engineering.

Interpolation and regression methods for computer aided engineering

Constructing a useful meta-model starting from a reduced number of simulations is by no means a trivial task. Mathematical and physical soundness, computational costs and prediction errors are not the only points to be taken into account when developing meta-models. Ergonomics of the software have to be considered in a wide sense. Engineers would like to grasp the general trends in the phenomena, especially when the behavior is nonlinear. Moreover, engineers would like to re-use the experience accumulated, in order to spread the possible advantages on different projects. When using meta-models, engineers should always keep in mind that this instrument allows a faster analysis than complex engineering models, however, interpolation and extrapolation introduce a new element of error that must be managed carefully.

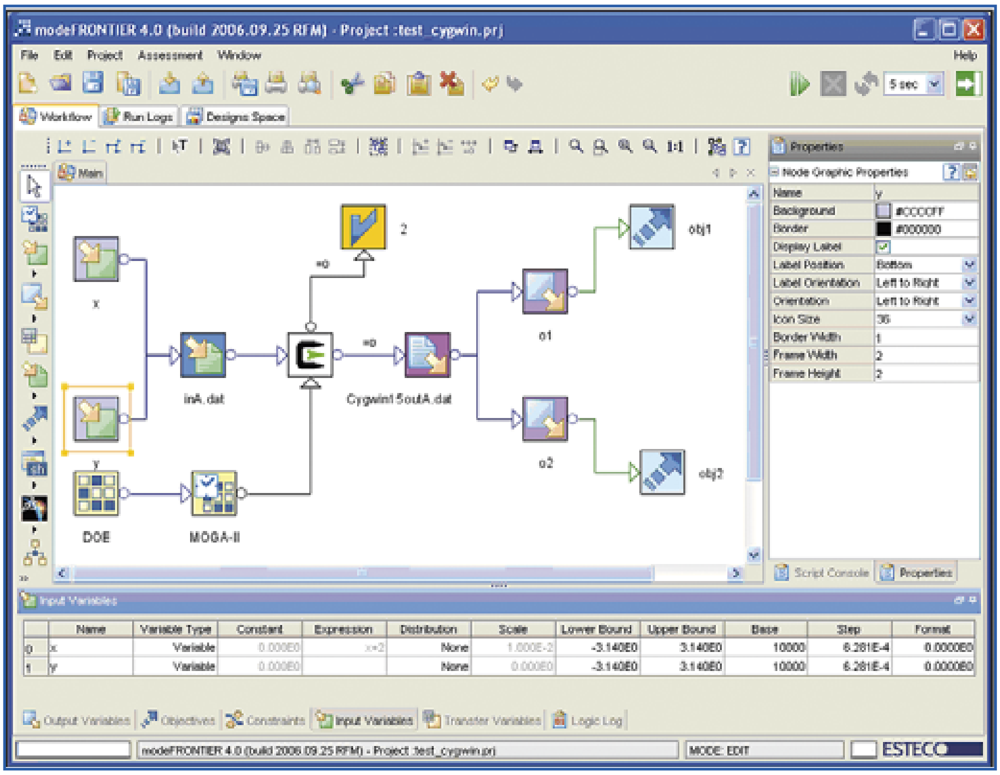

Figure 1: modeFRONTIER panel which helps engineers to easily formulate the problem, design the objectives and constraints, and identify the input and output parameters.

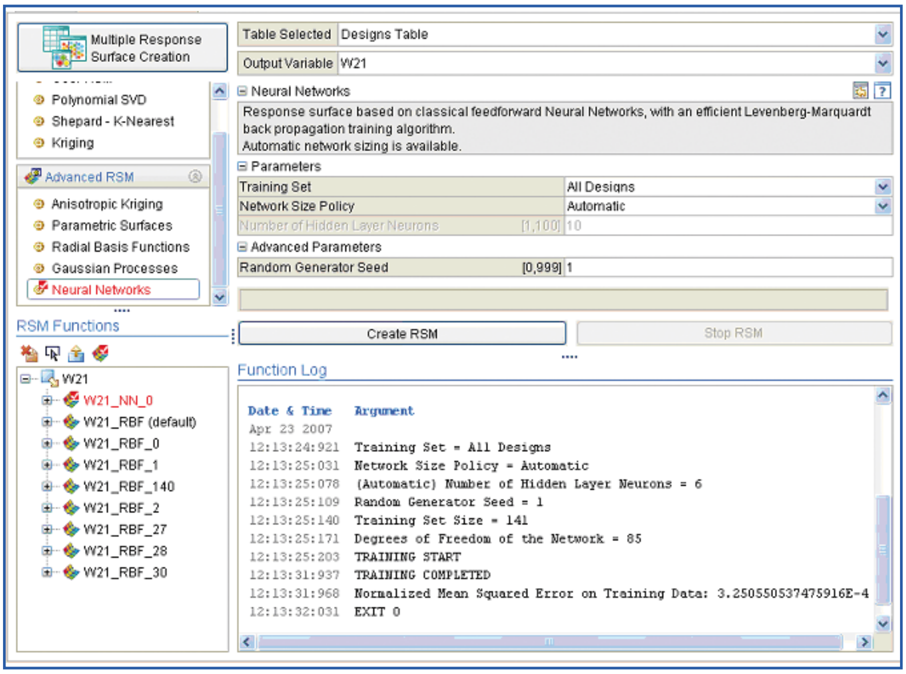

Figure 2: modeFRONTIER panel with which engineers can easily formulate, generate and save several kinds of meta-models.

It is for these reasons that in the last years, different approximation strategies have been developed to provide inexpensive meta-models of the simulation models to substitute computationally expensive modules. The intention of this article is to demonstrate particular features of modeFRONTIER that allow an easy use of the meta-modeling approach.

A typical sequence when using meta-models for engineering design can be summarized as follows:

1. First of all, engineers should formulate the problem, design the objectives and constraints, and identify the problem's input and output parameters; this may include specifying the names and bounds of the variables that will be part of the design, as well as characterizing the responses. This is quite an easy task in modeFRONTIER; the user can take advantage of all the features and the node of the workflow [fig.1]. At this step, it is also advisable to determine whether the use of a meta-model is justified, or whether the analysis should be conducted with the original simulation instead.

2. If the original simulation is computationally heavy and the use of the meta-model is necessary, the designer should choose the number and type of designs at which it would be more convenient to run the original simulation model. The true output responses obtained from these runs are used for fitting the meta-model. Even though this step is quite an easy task in modeFRONTIER, the user can take advantage of all the methodologies available in the Design of Experiment tool.

3. At this point, the engineer can use the output responses obtained in the previous step for fitting the meta-model. The fitting of meta-models requires specifying the type and functional form of the meta-model and the easy-to-use interface to save, evaluate and compare different responses. modeFRONTIER assists engineers even at this important step by means of its Response Surface Methodologies tools (RSMs) [Fig. 2].

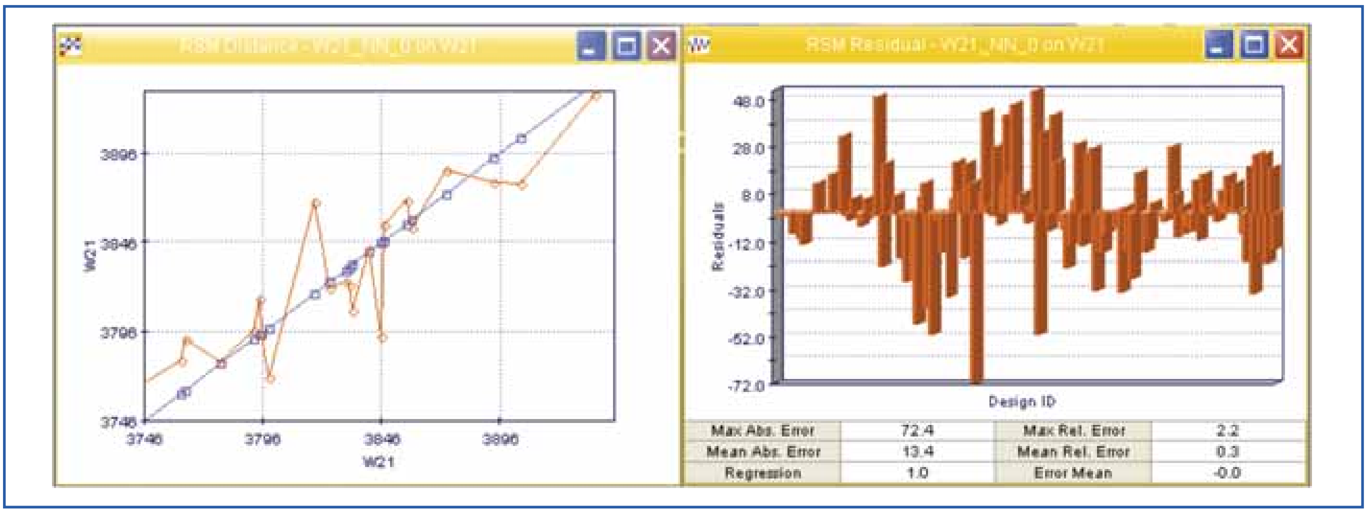

Figure 3: Distance chart (left) and residual chart (right). These charts represent two of the several possibilities offered by modeFRONTIER to validate the meta-models.

4. Another important step is the assessment of the meta-model which involves evaluating the performance of the models, as well as the choice of an appropriate validation strategy. In modeFRONTIER, the engineers have several charts and statistical tools at their disposal for evaluating the goodness of the meta-models [Fig. 3]. Gaining insight from the meta-model and its error permits the identification of important design variables and their effects on responses. This is necessary to understand the behavior of the model, to improve it or to redefine the region of interest in the design space.

5. The last step consists of the use of meta-models to predict responses at untried inputs and performing optimization runs, trade-off studies, or further exploring the design space. As these points are extracted from meta-models and not obtained through real simulations, they are considered virtual designs. Even this last step is quite an easy task in modeFRONTIER; the user can immediately re-use the generated meta-models to speed up the optimization.

Meta-models for laboratories

The previous section describes how meta-models can help to speed up optimization by substituting time consuming simulation models. A similar approach can be used to create synthetic models from experimental data.

In this case, the aim is to substitute a time consuming and probably costly experiment with a good enough mathematical model.

modeFRONTIER is able to import many file formats (XLS, TXT, CSV…), within a few easy steps. These designs resulting from experiments can be used to carry out statistical studies, such as sensitivity analysis, training for response surface modeling exactly as described in the previous sections.

In modeFRONTIER, all the tools for measuring the quality of meta-models in terms of statistical reliability are available. Moreover, modeFRONTIER gives a set of reasonable meta-modeling methods to interpolate different kinds of data. These methods include:

- Multivariate Polynomial Interpolation based on Singular Value Decomposition (SVD)

- K-Nearest, Shepard method and its generalizations. Shepard's method is a statistical interpolator which works through averaging the known values of the target function. The weights are assigned according to the reciprocal of the mutual distances between the target point and the training dataset points. The k–nearest method averages only on the most k nearest data to the target point. Shepard's method is one of the so called Point Schemes, i.e., interpolation methods which are not based on a tessellation of the underlying domain. Shepard is maybe the best known method among all scattered data interpolants in an arbitrary number of variables in which the interpolant assumes exactly the values of the data. The interpolated values are always constrained between the maximum and the minimum values of the points in the dataset. The response surface obtained with this method has a rather rough and coarse aspect, especially for small values of the exponent. Perhaps one of the most relevant drawbacks of this method is the lowering of maxima and the rising of minima. In fact, one usually expects that averaging methods like Shepard's flatten out

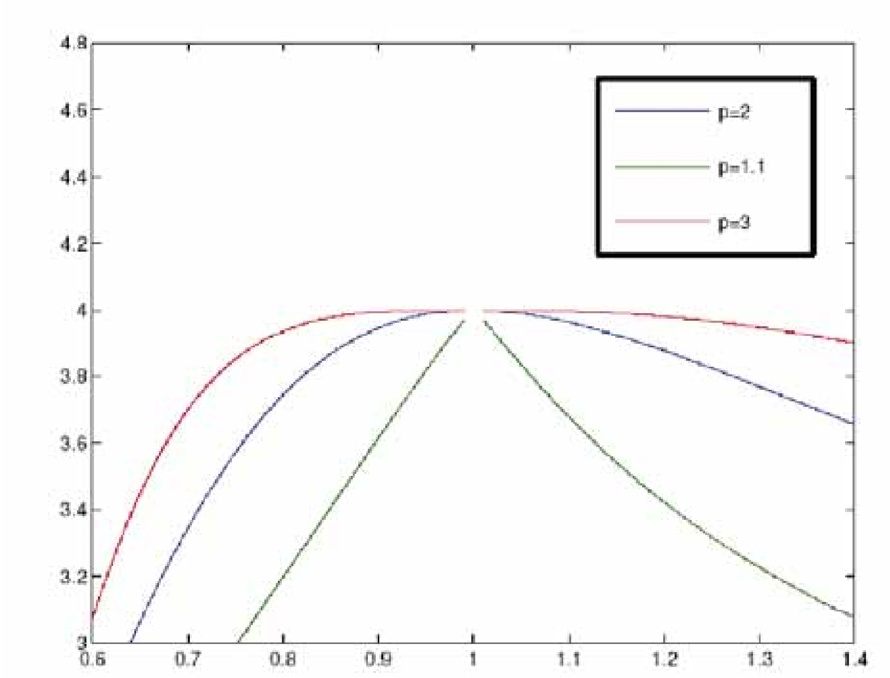

Figure 4: The effect of choosing different values of the characteristic exponent p in the weighting function around a point of the training dataset. The function flattens as long as the exponent grows.

extreme points. This property is particularly undesirable in the situation shown in Figure 4, where the interpolating model disastrously fails to describe the underlying function, which is an ordinary parabola. It is self–evident that this feature is crucial for seeking the extremes.

- Kriging: Kriging is a regression methodology that originated from the extensive work of Professor Daniel Krige, from the Witwatersrand University of South Africa, and especially from problems of gold extraction. The formalization and dissemination of this methodology, now universally employed in all branches of geostatistics, as oil extraction and idrology among others, is due to Professor Georges Matheron, who indicated the Krige's regression technique as krigeage. This is the reason why the pronunciation of kriging with a soft "g" seems to be the more correct one, despite the hard "g" pronunciation mainly diffused in the U.S. Thanks to the support of the Department of Mathematical Methods and Models for Scientific Applications of the University of Padova, modeFRONTIER contains a simple kriging featuring the four variogram models with the possibility of auto determination of the best fitting of the experimental variogram. The fitting procedure uses Levenberg–Marquardt to minimize the sum of the squares of the differences between values from the experimental variogram and values from the model. Moreover, the user is warned when the best fitting variogram shows some clue of unacceptability, or still larger than the larger differences between values in the dataset.

- Parametric Surfaces: Useful whenever the mathematical expression of the response is known, except for some unknown parameters. The training algorithm calculates the values of the unknown parameters that yield the best fit.

- Gaussian Processes: Implement the Bayesian approach to regression problems: The knowledge of the response is expressed in terms of probability distributions. This algorithm is best suited for non polynomial responses.

- Artificial Neural Networks: As well as many human inventions or technical devices, artificial Neural Networks take inspiration from Nature in order to realize a kind of calculator completely different from the classical Von Neumann machine, trying to implement at the same time the hardest features and tasks of computation as parallel computing, nonlinearity, adaptivity and self training. A neural network is a machine that is designed to model the way in which the brain performs a particular task or function of interest. To achieve their aims, neural networks massively employ mutual interconnections between simple computing cells usually called neurons. Networks simulate the brain in two aspects: The knowledge is acquired through a learning process and the information is stored in the synaptic weights, i.e., the strengths of the interconnections between neurons [Fig. 5]. The class of Neural Networks included in modeFRONTIER with a single hidden layer is shown to be capable to interpolate any functions with minimum request of regularity.

- Radial Basis Functions (available from version 4.0)

The description of each interpolation method constitutes by itself a separate topic and paper, hence going deeply into this kind of description is not the aim of this article.

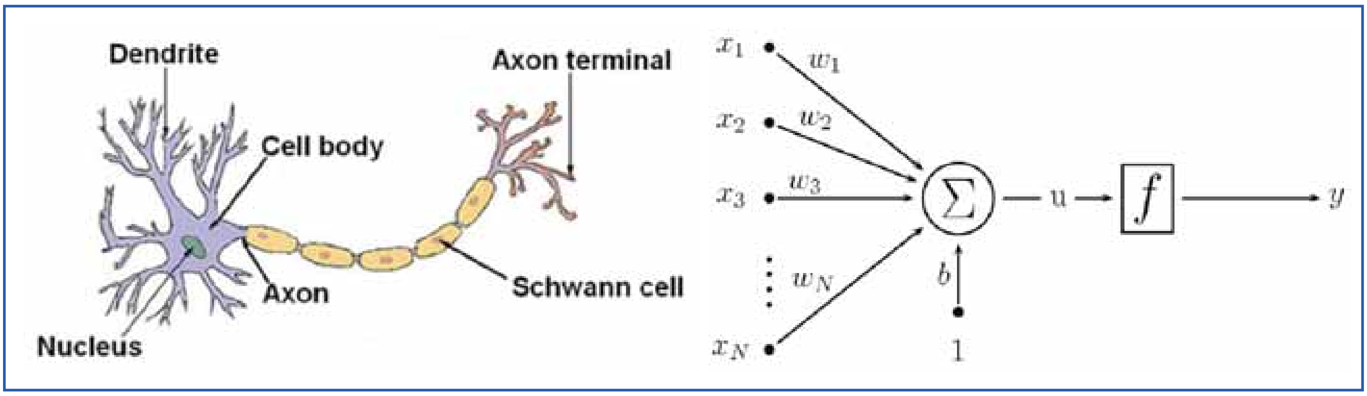

Figure 5: Neural Networks (NN) are inspired by the functioning of biological nervous systems. A real neuron (left) and an artificial neuron (right)

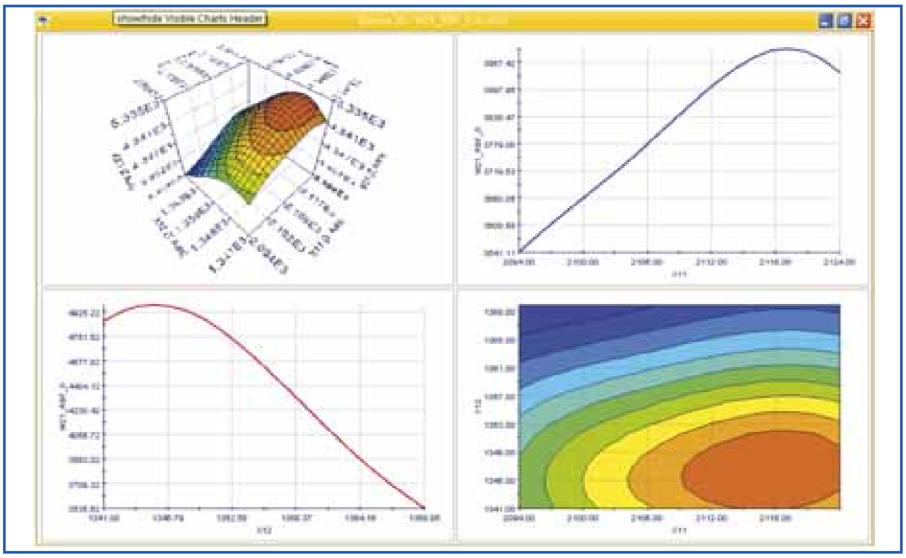

Figure 5: modeFRONTIER tool for meta-models 3D-exploration

Considering that several methods for interpolation are available, both in modeFRONTIER and in literature, an engineer may ask which is the best model to be used. There is an obvious notion that more simple functions can be approximated better and more complex functions are in general more difficult to approximate regardless of the meta-modeling type, design type and design size. A general recommendation is to use simple meta-models first (such as on low order polynomials). Kriging, Gaussian and Neural Network should be used for more complex responses. In general and regardless of the meta-model type, design type, or the complexity of the response, the performance tends to improve with the size of the design, especially for Kriging and Artificial Neural Networks.

Validation of Meta-models

modeFRONTIER has a powerful tool for the creation of meta-models, as it gives the possibility to verify the accuracy of a particular meta-model and to decide whether or not to improve its fidelity by adding additional simulation results to the database. It is possible to decide on effective surfaces for statistical analysis, for exploring candidate designs and for the use as surrogates in optimization. If the training points are not carefully chosen, the fitted model can be really poor and influence the final results. Inadequate approximations may lead to suboptimal designs or inefficient searches for optimal solutions.

That is why validation is a fundamental part of the modeling process. In modeFRONTIER, during the interpolation, a list of messages and errors generated by the algorithms is shown. The messages provide suggestions for a better tuning of the selected models. They generally list the maximum absolute error which is a measure that provides information about extreme performances of the model. The mean absolute error is the sum of the absolute errors divided by the number of data points, and is measured in the same units as the original data. The maximum absolute percent error is the maximum absolute numerical difference divided by the true value. The percentage error is this ratio expressed as a percent. The maximum absolute percent error provides a practical account of the error, measuring by what percentage a data point deviates from the mean error. There are many other measures that might be used for assessing the performance of a meta-model (e.g. the R-squared).

Conclusions

Both websites, www.esteco.com as well as www.network.modefrontier.eu, the portal of the European modeFRONTIER Network, provide several examples of how to use meta-modeling techniques to speed-up optimization.

For any questions on this article or to request further examples or information, please email the author:

Silvia Poles
ESTECO - Research Labs
scientific@esteco.com

References

[1] P. Alfeld. Scattered data interpolation in three or more variables. In T. Lyche and L. Schumaker, editors, Mathematical Methods in Computer Aided Geometric Design, pages 1–34. Academic Press, 1989.

[2] Martin D. Buhmann, Radial Basis Functions: Theory and Implementations Cambridge University Press, 2003

[3] Armin Iske, Multiresolution Methods in Scattered Data Modelling, Springer, 2004

[4] Georges Matheron. Les Variables Regionalisees et leur Estimation, une Application de la Theorie de Fonctions Aleatoires aux Sciences de la Nature. Masson et Cie, Paris, 1965.

[5] Holger Wendland, Scattered Data Approximation, Cambridge University Press, 2004