Energy Informatics

RESEARCH

Open Access

Surrogate models for composed simulation models in energy systems

Stephan Balduin

From The 7th DACH+ Conference on Energy Informatics
Oldenburg, Germany. 11-12 October 2018

Correspondence: stephan.balduin@
offis.de
OFFIS – Institute of Information
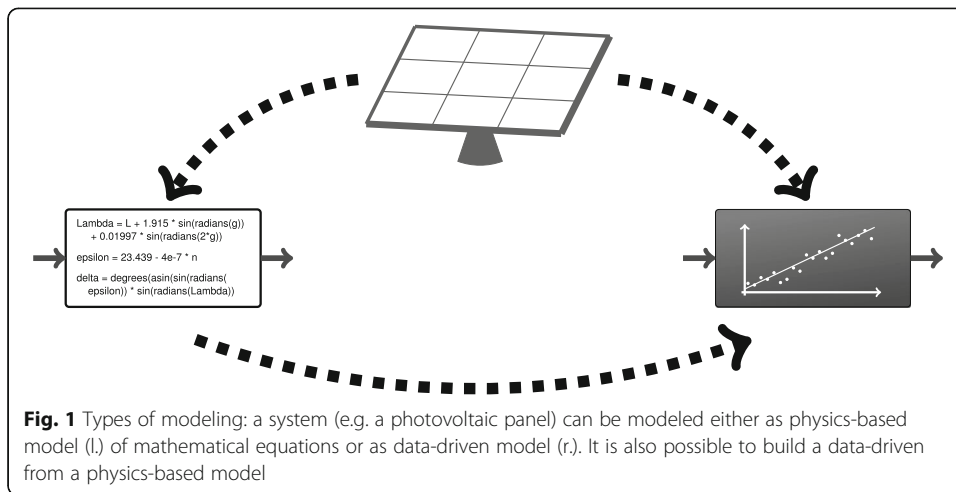Technology, Escherweg 2, 26121
Oldenburg, Germany

Abstract

New technologies and methodologies for smart grid applications cannot be tested in the real power grid, since it is a safety-critical infrastructure, therefore simulation and co-simulation is utilized. Simulation models itself can rely on quite complex calculations and therefore slow down the simulation. But even less complex models can lead to performance issues when used in large numbers in large-scale setups. The use of surrogate models is one way to improve the performance of simulation systems when the simulation models are slow, but the performance gain diminishes, when the simulation models are already quite fast. This abstract presents a new PhD project, which proposes a method to combine several simulation models into one surrogate model using correlations and other interdependencies of the simulation models. The goal is to further improve the performance gain not only for slower, but also for less complex simulation models, thus enable even larger simulation setups.

Keywords: Surrogate models, Machine learning, Simulation, Co-simulation, Energy system, Smart grid
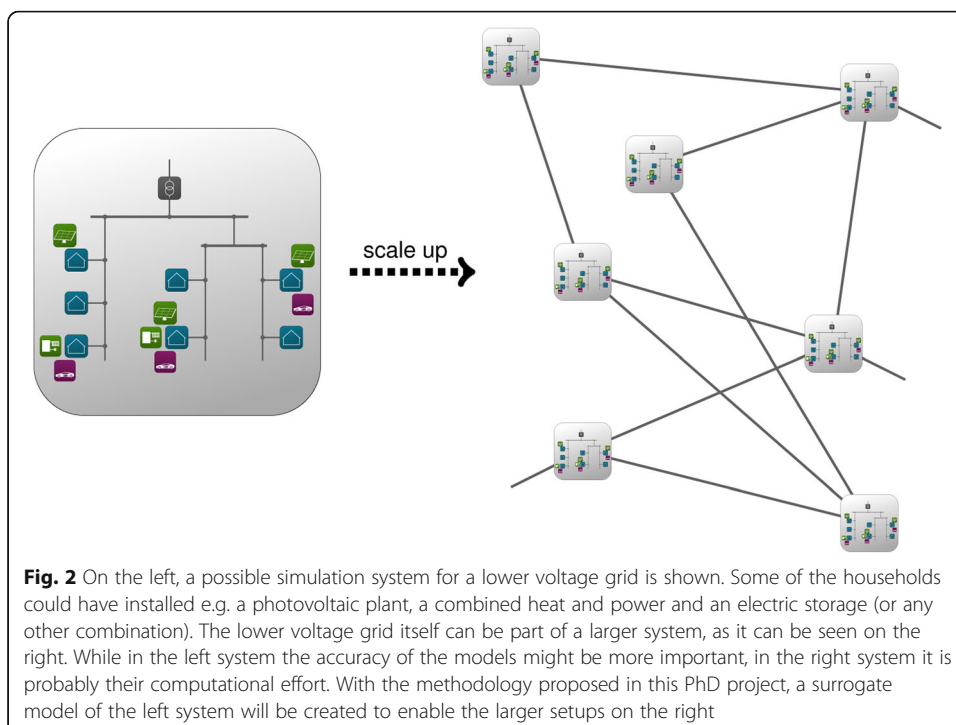
Background

The simulation and co-simulation of the power grid has been established as a tool for testing and assessing new technologies and methodologies, as the real power grid as a safety-critical infrastructure is not suitable for testing (Nieße et al., 2014). The models used in such simulations are only abstractions of their underlying systems, but can also be quite complex and therefore expensive to compute. While this is negligible in small simulation systems, even simpler models can slow down the simulation for larger setups. In various applications fast approximation models have been used as suitable replacement for complex simulation models to tackle these issues. Such approximation functions are also called response surfaces or surrogate models and are built from sample data, i.e. inputs for the original system and corresponding output values (see Fig. 1). Currently, surrogate models in energy system simulations are often designed to treat the simulation model as a black box, and usually there are good reasons for doing so, e.g., because the simulation model was designed by an domain expert and is therefore too complex for the surrogate modeler to understand. The modeler would have to invest time to familiarize himself with the domain in question, which is not feasible in

Fig. 1 Types of modeling: a system (e.g. a photovoltaic panel) can be modeled either as physics-based model (l.) of mathematical equations or as data-driven model (r.). It is also possible to build a data-driven from a physics-based model

a setup with multiple simulation models from different domains. On the one hand, smart grid simulation models are often computer models where machine learning is generally well-suited to create surrogates from black box simulation models. On the other hand, less attention has been paid to the simulation models treated as gray or white boxes. This may be because the need for internal knowledge about a model becomes less and less important for the surrogate modeler since only input and output data is required. However, in large-scale simulation systems (see Fig. 2), it may not be sufficient to substitute surrogates for individual simulation models.

It might be better if several of these simulation models were aggregated in one surrogate model in order to reduce the computing time that would otherwise be necessary for the calculation of the individual simulation models. Furthermore, in

Fig. 2 On the left, a possible simulation system for a lower voltage grid is shown. Some of the households could have installed e.g. a photovoltaic plant, a combined heat and power and an electric storage (or any other combination). The lower voltage grid itself can be part of a larger system, as it can be seen on the right. While in the left system the accuracy of the models might be more important, in the right system it is probably their computational effort. With the methodology proposed in this PhD project, a surrogate model of the left system will be created to enable the larger setups on the right

large-scale setups the exact behavior of individual models is less important than the interaction of larger numbers of components, but it is still unclear whether the consideration as black box provides satisfactory results. To this end, this PhD project attempts to provide a methodology for combining multiple simulation models into one surrogate model and to evaluate them against the simulation models as well as against a pure black box model.

Research question

The goal of the PhD project presented in this abstract is to define a process for building surrogate models from complex or aggregated simulation models for large-scale energy system simulations. These surrogates are not intended to be simple black boxes, but use additional information gained from simulation experiments or knowledge about the system structures. This leads to the top-level research question:

(RQ) How can the aggregation of multiple simulation models into one surrogate model to improve the performance in co-simulation setups be realized?

This may also include single but complex simulation models as well as the treatment of simulation models as black boxes using additional information, such as correlations of internal states, output behavior over time, other shared patterns, or in general, their relationship within the simulation derived from knowledge about the setup. The correlations in particular are of interest for the aggregation and an appropriate correlation model needs to be selected. Thus, the first subquestion will be:

(RQ 1) Which correlations and dependencies exist between the models and how can they be used for the aggregation?

A database is needed, which can be derived from experimental data during so-called screening experiments as well as from other sources, such as the system design. Especially these other sources need to be identified, as well as what kind of information can be obtained from simulation experiments. Another aspect could be the effort required to improve the model, e.g. if simulation runs take a long time or are expensive and therefore the amount of training has to be kept small. Further, it is essential to identify and evaluate different aggregation strategies in order to know, which kind of information is required. This can be comprised in the second subquestion:

(RQ 2) What kind of information about the models and the simulation system is needed and how can it be used to aggregate simulation models?

After all these information is available, the next step will be the training process of the surrogate model. A detailed strategy is needed for the composition of the gathered information to build a training data set. It is also possible that additional sampling is required and the training algorithm should be appropriately selected to work with the available information and the aggregation strategy. Therefore, the third subquestion is:

(**RQ 3**) What are possible strategies to construct a surrogate model incorporating specified information and dependencies?

A metric is required to evaluate the quality of the trained model and several aspects should be considered. There is no benefit of the surrogate model, if it performs equal or even worse than the original model. Thus one part of the metric should evaluate the performance of the surrogate model. This is addressed by the fourth subquestion:

(**RQ 4**) How much benefit brings the usage of surrogate models instead of the simulation models in a co-simulation setup?

Since simulation models represent abstractions that are subject to errors or uncertainty, and surrogate models in turn represent abstractions of the simulation models, this uncertainty increases even further. Therefore, another part of the metric should quantify the uncertainty of the surrogate model overall and in respect to the simulation model. This leads to the fifth subquestion:

(**RQ 5**) How much impact has the use of surrogate models on the uncertainty of the simulation?

Related work

Currently, five relevant research topics have been addressed: surrogate modeling, energy system simulation, correlation modeling, model composition and uncertainty analysis. State of the art and related work regarding these topics will be presented in this section.

The surrogate modeling has its origins in and is part of statistical design of experiments, which is a domain-independent tool for describing observed behavior of a system. There is a lot of literature on the subject and for a deeper insight the reader is referred to Response Surface Methodology of Myers et al. (Myers et al., 2016), which represents a comprehensive state of the art. A variety of practical applications of surrogate models can be found. In the research project D-Flex,[1] the integration of renewable energy resources based on a centrally controlled load and generation management approach was compared to a decentralized approach. The evaluation was model-based and carried out as part of a scenario analysis. To estimate the required computation time, a benchmark scenario with 70,000 units to be simulated was defined and run for a simulated time of 1 day, 1 week and 1 month. The latter took almost 5 days to complete. Surrogate models successfully reduced the calculation time and thus made it possible to carry out the planned evaluations with several scenarios for 1 year each within time. Dalal et al. investigated outage scheduling for components of the power system (Dalal et al., 2018). Outage scheduling is necessary to organize maintenance and replacement activities of components. They presented a framework to assess outage schedules and proposed an optimization method to create a schedule for a list of required outages. In outage scheduling, several future scenarios are evaluated in terms of feasibility. The authors assessed the evaluation of a large number of possible scenarios as impracticable and therefore machine learning was utilized to generate a surrogate model that evaluates these scenarios.

An increasing number of distributed energy resources means that not only a few large energy generators have to be controlled, but also many small ones. In addition, there are also consumers whose consumption can be controlled in terms of time or quantity. The coordination of all these units requires a functioning communication network. The combination of the power grid and information and communication technology is called smart grid. Since testing technologies and methodologies in the real power grid is not feasible, simulation is the tool of choice. The development of technologies and algorithms for smart grids can practically not be carried out in the real power grid. Steinbrink et al. (Steinbrink et al., 2017) gave an overview over the state-of-the-art simulation-based approaches, which is summarized in the following. To simulate the individual components of the smart grid, simulation models of these components are required, some of which can be very complex. These models are often built by domain experts for their favorite simulation environment. To couple all these simulation models and environments, co-simulation can be used, i.e. each simulator only needs to implement the interface for the co-simulation framework, which handles communication among different simulators. Steinbrink et al. conclude that co-simulation is one of three suitable tools for smart grid simulation. They also point out, that future research needs to include the development of surrogate models to improve simulation performance. The other two simulation tools, multi domain simulation as well as real time simulation and hardware-in-the-loop, will not be discussed here.

To model dependencies between two random variables, different correlation functions can be used. Linear dependency is often described with the Pearson correlation, which returns a value between − 1 and 1. Positive correlation implies, that both variables simultaneoulsy attain high or low values. On the other hand, negative correlation implies: When the first variable attains a high value, the second will attain a low value. But it is also possible to have non-linear dependencies and the Pearson correlation is too restricted to model these. Such dependencies of two or more random variables can be fully described with multivariate or joint distribution functions. When high numbers of random variables with dependencies are expected, a partial correlation network can be useful. Partial correlation is the dependency between two random variables without the influence of other variables. A network of partial correlation visualizes the dependencies and is, e.g., used in psychological science (Epskamp & Fried, 2018).

The idea of model composition is basically to utilize dependencies between two random variables, which can be from the same model or from other models. Blank proposed a method to assess the reliability of coalitions of renewable power units for the provision of ancillary services in her PhD thesis (Blank, 2015). The composition of such a coalition is connected to the planning of how much power the units produce and when. The units considered by Blank are wind and photovoltaic systems that are located spatially close to each other. This is relevant for risk analysis, as it can be assumed that dependencies exist between the forecasting errors of the units. These dependencies can be modeled by correlations or, in non-linear cases, by joint distributions. Another approach for model composition is the so-called cokriging method. Han et al. (Han et al., 2010) adapted the general idea of cokriging and used it for variable fidelity modeling, i.e. combining two datasets describing the same model. The assumption is, that one of these datasets comes from a high-fidelity model, which is expensive to compute and therefore this dataset contains not very many samples.

The other dataset is produced by a low-fidelity model and has much more samples than the first one. Cokriging interpolates between these two datasets and therefore aggregates the low-fidelity and the high-fidelity model.

In his PhD thesis, Steinbrink (Strinbrink, 2017) developed a modular concept for uncertainty quantification in smart grid co-simulation. A prototypical implementation for the co-simulation framework mosaik[2] was also provided with certain energy simulation systems in different sizes. To quantify the uncertainty of a model, sample data from simulation steps is needed and higher number of samples can improve the accuracy of the uncertainty quantification. The author also utilizes simple interpolation models as surrogates to reduce the number of required samples compared to a Monte Carlo sampling approach. Wilson et al. (Wilson et al., 2018) investigated a computer model of the UK's electricity supply with regard to uncertainty. This model calculates electricity price projections from 2010 to 2030 and uses uncertain inputs, such as future energy demand. The uncertainty of the computer model caused by these and other influencing factors should be quantified. Since the evaluation of the model took up to 1 hour, the authors used a Bayesian linear model as a surrogate model. In this way, the number of necessary computer model evaluations could be greatly reduced, even if this added another source of uncertainty. This surrogate model was used together with a probability distribution over the inputs to study the uncertainty of the overall model. The authors conclude that surrogate models are a useful approach for the quantification of uncertainties, especially if the number of evaluations of the original model is to be kept low for time reasons.

Surrogate models are an established tool to speed-up simulations and there are also many applications in the energy domain, that make use of surrogate models. There is also work done that addresses the composition of different units or models by determining interdependencies and correlations. But there is still missing a methodology to use these dependency information in order to build a surrogate model, that comprises two or more simulation models. The contribution of this PhD project is to link these topics and to derive a methodology which closes this gap.

Methodology

The research method of this PhD project is based on Design Science by Peffers et al. (Peffers et al., 2007). The problem identification and motivation was described in the background section of this abstract. The objectives for the solution this project proposes were informally described in the research question section and still need further refinement. In order to answer the research question and subquestions, an environment is needed in which simulation models can be investigated and surrogate models can be trained and tested. Before the design and development process can be started, it is necessary to define such an environment in a preparation step that can be later used in the demonstration and evaluation steps. It may be useful to identify two or, better, more different systems within this environment not only to have a testbed for development and evaluation, but also to estimate, how well the methodology performs on different problems. The first step of the proposed methodology will involve an investigation of cross-model dependencies (RQ1), i.e. correlations and partial correlations. This is followed by a comprehensive analysis of the simulation models (RQ2), i.e. inputs, outputs, response properties and interactions, which is, in the surrogate modeling

literature, also known as sensitivity analysis or screening (Simpson et al., 2001). Further, an aggregation strategy should be selected and applied. The results from the first step can be used to select experimental designs from which samples can then be generated as well as appropriate surrogate models to represent the simulation models (RQ3). For a suitable selection it is necessary to know how expensive a simulation run is, i.e. how long the machine runs to produce a sample. Since the energy simulation is performed using computer-based models, it can be assumed that samples can be easily generated and therefore one of the many available space-filling designs for computer experiments can be selected. If this is not the case, the samples should be selected more strategically and therefore one of the methods of the classical response surface methodology could fit. It should be borne in mind that certain surrogate methods are likely to work better with certain classes of sampling designs and it is also necessary to know how many samples are feasible to generate. More samples mean a wider range of possible methods, e.g. artificial neural networks which can take a large numbers of samples. The last step is to evaluate the model and research subquestions four and five point in two different directions. The first aspect is the benefit of the surrogate model, i.e. the performance compared to the simulation models. This requires, e.g., a runtime analysis of both models within the simulation systems defined in the preparation step. The second aspect is the impact of the surrogate model on the uncertainty of the simulation system, i.e. not only a comparison with the replaced simulation model is considered, but also the other components of the simulation setup.

Conclusion

In this abstract, a new PhD project was presented. The idea and the goal of this project is to provide a method to aggregate multiple simulation models into one surrogate model and evaluate it against performance, accuracy and uncertainty. These surrogates serve in a co-simulation system as fast approximation and therefore enable the creation of larger setups. Since this project is in an early stage, research goals and methodology are quite abstract yet and need further literature research and refinement, which will be done in the next steps. Furthermore, appropriate environments will be identified for the preparation step.

Endnotes

[1]D-Flex Final Report [https://www.offis.de/offis/projekt/d-flex.html].
[2]Mosaik Website [https://mosaik.offis.de].

About this Supplement
This article has been published as part of *Energy Informatics* Volume 1 Supplement 1, 2018: Proceedings of the 7th DACH+ Conference on Energy Informatics. The full contents of the supplement are available online at https://energyinformatics.springeropen.com/articles/supplements/volume-1-supplement-1.

Author's contributions

The author has read and approved the final manuscript. The content of the manuscript was created by author SB, unless otherwise indicated.

Competing interests

The author declares that he has no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Blank D-MM (2015) Reliability assessment of coalitions for the provision of ancillary services. University of Oldenburg

Dalal G, Gilboa E, Mannor S, Wehenkel L (2018) Chance-Constrained Outage Scheduling using a Machine Learning Proxy. arXiv preprint arXiv 1801:00500

Epskamp S, Fried EI (2018) A tutorial on regularized partial correlation networks. Psychol Methods

Han Z-H, Zimmermann R, Goretz S (2010) A new cokriging method for variable-fidelity surrogate modeling of aerodynamic data. In: 48th AIAA Aerospace sciences meeting including the new horizons forum and Aerospace exposition

Myers RH, Montgomery DC, Anderson-Cook CM (2016) Response surface methodology: process and product optimization using designed experiments. John Wiley & Sons

Nieße A, Tröschel M, Sonnensche M (2014) Designing dependable and sustainable smart grids - how to apply algorithm engineering to distributed control in power systems. Environ Model Softw 56:37–51

Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. J Manag Inf Syst 24:3

Simpson TW, Poplinski JD, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: survey and recommendations. Eng Comput 17:2

Steinbrink C, Lehnhoff S, Rohjans S, Strasser T, Widl IE, Moyo C, Lauss G, Lehfuss F, Faschang M, Palensky P et al (2017) Simulation-based validation of smart grids - status quo and future research trends. In: International Conference on Industrial Applications of Holonic and Multi-Agent Systems

Strinbrink C (2017) A nonintrusive uncertainty quantification system for modular smart grid co-simulation. University of Oldenburg

Wilson AL, Dent CJ, Goldstein M (2018) Quantifying uncertainty in wholesale electricity price projections using Bayesian emulation of a generation investment model. In: Sustainable Energy, Grids and Networks, p 13