Energy Informatics

RESEARCH

Open Access

Evaluating different machine learning techniques as surrogate for low voltage grids

Stephan Balduin[1*], Tom Westermann[2] and Erika Puiutta[1]

Correspondence:
stephan.balduin@offis.de
[1]OFFIS – Institute for Information
Technology, Escherweg 2, 26121,
Oldenburg DE, Germany
Full list of author information is
available at the end of the article

Abstract

The transition of the power grid requires new technologies and methodologies, which can only be developed and tested in simulations. Especially larger simulation setups with many levels of detail can become quite slow. Therefore, the number of possible simulation evaluations decreases. One solution to overcome this issue is to use surrogate models, i. e., data-driven approximations of (sub)systems. In a recent work, we built a surrogate model for a low voltage grid using artificial neural networks, which achieved satisfying results. However, there were still open questions regarding the assumptions and simplifications made. In this paper, we present the results of our ongoing research, which answer some of these questions. We compare different machine learning algorithms as surrogate models and exchange the grid topology and size. In a set of experiments, we show that algorithms based on linear regression and artificial neural networks yield the best results independent of the grid topology. Furthermore, adding volatile energy generation and a variable phase angle does not decrease the quality of the surrogate models.

Keywords: Machine learning, Artificial neural network, Surrogate model, Power grid, Power flow

Introduction

The ongoing transformationof the power system requires the involvement of new technologies and methodologies to meet the requirements that arise during this process. Since the power grid is a safety-critical infrastructure, simulation and hardware-in-the-loop are used for the development and testing of such new technologies (Steinbrink et al. 2017). Smart grid simulations comprise also other domains, such as the Information and Communication Technology (ICT) domain, each of which are developed in their own simulation environment. With co-simulation, synchronization and data exchange between these different environments is handled by a co-simulation-framework (e. g., the co-simulation framework mosaik[1]).

---

[1]https://mosaik.offis.de, retrieved on 10 Jun. 2020

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 2 of 12

Building such a large, cross-domain simulation environment is still a complex task and the simulation of the overall system can become very slow (Blank et al. 2015). To overcome this, surrogate models can be used to reduce the simulation time of some of the components in the environment. Usually, a surrogate model is a data-driven approximation of a certain function or system, which can be evaluated faster than the original function or system, but which is less accurate (Simpson et al. 2001).

In a recent work, we built a surrogate model for a subsystem of such a multi-domain co-simulation environment (Balduin et al. 2019). Starting with the CIGRE low voltage (LV) benchmark power grid model (Papathanassiou et al. 2005), we added several load models that were connected to certain buses in the grid and built a surrogate model to replace these components. We evaluated the surrogate model in a case study with several experiments. Our goal was to enable the creation of larger setups of medium voltage (MV) grid simulations using this kind of surrogate model as a replacement for some of the LV grids.

In this paper, we aim to address some of the open issues of that work. In the next section, we give an overview of "Related work" done in this field. This is followed by a description of the "Simulation setup" created in the previous work, which we modified for this work. In the subsequent section, we present the "Case study" of this work, comprising hypotheses, experiments, and the presentation of the results. We conclude our work and discuss the results in the "Conclusion". Finally, we discuss open questions and future work in the "Outlook" sections. Since this article had to be shortened to a short paper for publication, some information might be presented incomplete. An extended preprint version of this paper can be found in (Balduin et al. 2020).

Related work

Surrogate models are a common technique from the field of statistical design of experiments (Myers et al. 2016). They are used to describe the behavior of a system that, for various reasons, is not suited to be built knowledge-based. They are used in a broad range of use cases in the energy domain: starting from the calculation and optimization of energy savings (Beisheim et al. 2019; Nagpal et al. 2019; Vazquez-Canteli et al. 2019) and the replacement of specific simulation models (Balduin et al. 2019; Dimitrov 2019) over surrogate models for (micro)grids (Baumann et al. 2019; Balduin 2018; Grundel et al. 2019) to the use in uncertainty and reliability assessment (Blank and Lehnhoff 2014; Slot et al. 2020; Steinbrink 2016). This list is far from complete and there are also other approaches such as in Gerster (2018) who use surrogate models to build a decoder function abstracting from technical system specifications.

When used for grid emulation, surrogate models are basically used to replace a power flow (PF) analysis and, in this paper, the PF analysis assumes that the power grid is in a steady-state. The most common traditional methods for PF are Newton-Raphson (NR), Gauss-Seidel (GS), and derivatives, which calculate bus voltages, currents, etc. numerically. These methods have some important drawbacks. Most of them require to perform matrix inversions, which are solved iteratively. Bad initial guessing of unknown values in this process can lead to divergence and, subsequently, a repetition is required. Therefore, other solutions to this problem are actively researched (e. g., in Grisales-Noreña et al. (2020), Kontis et al. (2019), Montoya et al. (2020), and Yuan et al. (2019)).

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 3 of 12

There is a broad variety of approaches to solve the PF and related problems with data-driven algorithms. Nilsson et al. (2018) built a ML-based simplified grid model to perform a PF analysis. The authors concluded that their model was good enough to be used in several applications such as security-constraint dispatch and intra-hour simulations. Frank et al. (2012) surveyed about data-driven approaches to the closely related optimal power flow (OPF) problem. More approaches can be found in (Donon et al. 2020; Gupta et al. 2015; Syai'in and Soeprijanto 2010; Veerasamy et al. 2020; Xiang et al. 2020).

One question when applying ML to power grid models is related to the gathering of training data. Danner and de Meer (2019) discussed several methods to perform state estimation in the LV distribution grid. To solve the sampling problem, the authors compared Monte-Carlo simulation with one-at-a-time sensitivity analysis and realistic load profiles. Although they did not present final results, their proposed methodology sounds interesting and worthy to investigate further.

A quite different approach is given by Zhao et al. (2019) who built a parametric surrogate model. The authors showed that such a parametric surrogate model can reduce the dimensionality of the model and allows to embed physical constraints, which were required for their case studies. As a further benefit, Zhao et al. argued that their model could also improve the interpretability of the trained models. Interpretability and explainability are a current issue when it comes to ML models that are used in critical decision making, especially for deep learning (DL) and reinforcement learning (RL). For a deeper insight in the field of explainability of AI and RL we refer to (Puiutta and Veith 2020).

Another possible approach would be the use of RL to explore the grid and to aid in the sampling process. The concept of adversarial resilience learning (ARL) (Fischer et al. 2019) utilizes two classes of RL agents competing in a shared environment (Veith et al. 2020). The ARL methodology could be used to explore the sampling space, localize limitations of the model, and identify critical spots to derive sampling rules.

Simulation setup

The simulation setup in Balduin et al. (2019) used the pandapower (Thurner et al. 2018) implementation of the CIGRE LV benchmark grid (Papathanassiou et al. 2005). This grid consisted of three subgrids: a residential subgrid, an industrial subgrid, and a commercial subgrid. In each step, the load models forwarded the corresponding value from a time series to the grid. More information about the data sets can be found in Balduin et al. (2019). The load models and the pandapower grid model were coupled using the co-simulation framework mosaik. This assignment from load model to grid node was considered as part of the domain knowledge about the grid, which the surrogate model had to learn. Using unassigned load data, the surrogate model should predict the (vm_pu) of the grid's buses.

For this work, we included ML models with different characteristics, which we describe further in "Model selection" section All models were compared in experiments similar to the experiments in the reference work. The second goal of this work was to evaluate this methodology on a different simulation setup. The simbench project (Spalthoff et al. 2019) provides data sets for power grid benchmarks, including certain grid topologies. The *LV-rural3* data set describes a LV grid including time series of household loads and PV power generation. The simbench grids are available for pandapower which enabled us to extend the architecture without much additional work.

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 4 of 12

Beforehand, we analyzed and compared both data sets. Each time series contains 15-minute averages over one year resulting in a total of 35,400 entries. The residential loads tend to be higher in the cold months, lower in summer, and have noticeable fluctuations. A more regular behavior can be seen at the commercial and industrial loads (only CIGRE LV), which, unlike the residential loads, tend to be higher in summer. The PV generation (only LV-rural3) is higher in summer, remarkably low in December and January, and heavily depending on the weather conditions.

Finally, we compared the different topologies. The CIGRE LV grid includes 44 buses and 15 loads. These are distributed over the three subgrids mentioned before. The LV-rural3 has considerably more components: 128 buses, 118 loads, and 17 PV plants. In contrast to the CIGRE grid, LV-rural3 includes time series for active and reactive power, providing a varying phase angle.

Case study

Based on this setup, we conducted a case study to validate our extensions against the open questions of the previous work.

Hypotheses

The work this study is based on, aimed to provide a benchmark model and evaluation environment for further experiments as a proof-of-concept. This explains why (a) the architecture of the ANN was rather simple and (b) there was no reasoned selection of ANNs as a surrogate model in general. We aimed to provide a well-founded basis for these crucial points, selected a variety of models, and improved the hyperparameter tuning (see "Model selection"). Hereafter, any mention of a part of the study (e. g., hypothesis, or surrogate model) is referring to the corresponding part in the reference (Balduin et al. 2019). To compare these new models to the reference model, we defined three hypotheses similar to the reference hypotheses. The first one deals with the prediction quality of the models. To be predicted were the vm_pu of the buses, which were normalized values that should lie in the interval [0.9, 1.1]. We define a quality threshold of 0.1% for the prediction error (see "Description of experiment 1" sections for the reasoning).

Hypothesis 1: Surrogate Model Accuracy The behaviour of the voltage magnitudes of a low-voltage power grid simulation model can be adequately captured by a surrogate model (H_0: error of the models > 0.1%, H_a: error of the models ≤ 0.1%).

The main purpose of using surrogate models is to reduce the complexity of the original system. Therefore, the second hypothesis dealt with the question of whether the surrogate models actually reduced the computation time required for the simulation. For this experiment, we will defined an arbitrary time frame that was the same for all models. Each model repeatedly had to calculate the values for this time frame. The averaged results were compared to each other.

Hypothesis 2: Calculation Speed The surrogate models' calculation time t_{sur} on a defined time frame differs significantly from the simulation models' calculation time t_{sim} (H_0: $t_{sur} = t_{sim}$ ($\alpha \geq 0.05$), H_a: $t_{sur} \neq t_{sim}$ ($\alpha < 0.05$)).

In the third hypothesis, we verified whether our findings can be transferred to other grid topologies as well. With LV-rural3, some of the simplifying assumptions were eliminated and one step toward a generalizability check was done. However, since both grids are too different to be compared quantitatively, we performed a qualitative comparison.

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 5 of 12

Hypothesis 3: Generalization The results concerning the surrogate models' prediction accuracy and their calculation speed can be generalized to other simulation models (no test criterion, see "Description of experiment 2" section).

Model selection

Our goal was to compare models with different characteristics to provide an overview of which class of models performed best for the given task. In addition to the reference model, we selected models from four different families of machine learning algorithms.

First, we distinguished between single-target and multi-target models. Linear regression (LR) belongs to the former class and requires a fairly low number of calculation steps. Random forests (RF) are an ensemble method of decision trees and are more flexible than regular LR due to it's ability to also model non-linear input-output relationships. However, large numbers of trees slow down the training process. To provide multiple outputs as required by the given task, both models were combined to a regressor ensemble (RE). Additionally, LR was evaluated in a regressor chain (RC), which is mathematically equivalent, but is potentially more efficient to compute (Spyromitros-Xioufis et al. 2016). In total, three single-target models are evaluated – regressor ensemble linear regression (RE LR), regressor ensemble random forest (RE RF), and regressor chain linear regression (RC LR) – and their hyperparameters were optimized with random-search cross-validation.

We also selected three multi-target models. As a distance-based model we chose k-nearest neighbors (k-NN). K-NN has only a few hyperparameters with *k*, the number of neighbors to be considered, being the most prominent one. These will be optimized with a grid-search cross-validation. The second multi-target model we selected was the long short-term memory (LSTM) network, which is an adaption of ANNs specifically suited to temporal data (Hochreiter and Schmidhuber 1997). These kinds of networks are able to consider past values in the prediction and have a large number of hyperparameters, thus we decided to use the *hyperopt* hyperparameter optimization. Finally, we selected the same kind of ANN that was used in the reference work but with a more fine-grained architecture design. In contrast to the reference model, which consisted solely of fully-connected layers, we added task-specific layers, which forward their activation only to few or even only to one subsequent node. Furthermore, we inserted dropout layers after each hidden layer. In addition to number of epochs and the number of hidden layers, which were the only hyperparameters that were optimized in the reference model, we added batch size, different activation functions, a dropout factor, the number of task-specific layers, and the learning rate of the ANNs optimizer as hyperparameters. With these six models, we were able to make a reasonable assessment as to which models were best suited for the given use case.

Experimental setup

Description of experiment 1

The goal of the first experiment was to verify whether the surrogate models were able to accurately predict the vm_pus calculated by the simulation model. The load time series consisted of twelve months of data, of which eleven were used for training and one for evaluation. Since the data set showed significant differences between different months, the process was repeated twelve times in the form of a 12-fold cross validation. To quantify

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 6 of 12

the prediction error, the average root mean squared error (RMSE) over all n observations and all m bus vm_pus of the LV grid was calculated.

Since the voltage magnitudes were given in the per-unit-system, the observed ranges of voltage magnitudes spans from 0.87 to 1.0 for CIGRE LV and from 0.95 to 1.05 for the LV-rural3. For this reason, we chose to set the threshold for an adequate prediction error at an RMSE of 10^{-3} and skipped further normalization of the error. This corresponded to approximately one percent of the observed values' range. Therefore, we could reject the null hypothesis for a given surrogate model, if the average RMSE over the span of a year is greater than 10^{-3}. In order to also be able to judge whether or not the obtained results were robust towards changes in the LV grids' parameters, the experiment was repeated on the LV-rural3 grid.

Description of experiment 2

The second experiment aimed to provide insight into the surrogate models' calculation speed when used in a co-simulation-setup. Each of the models we built in experiment 1 was part of a separate co-simulation setup and had to calculate all voltage magnitudes over the span of the entire data set of 365 days. To mitigate the effect of process scheduling during the experiment, the calculation was repeated numerous times. In order to check the generalization of the benchmark's results to a different simulation model, we experimented with both the CIGRE LV and LV-rural3 simulation models. For the CIGRE LV, $n = 10$ repetitions were performed. Since the calculation time of the LV-rural3 was substantially longer than that of the CIGRE LV, only $n = 3$ repetitions were performed. For every single one of those n independent simulation-runs the calculation time was logged, so that we could compute the mean calculation time and its variance.

We conducted an analysis of variance (ANOVA) to test whether the differences between the results of the models were significant. To verify the assumptions for the ANOVA, we started to test for homogeneity of variances with Levene's test, which was not significant for neither of grids, i. e., with $F(6, 63) = 2.12, p \geq 0.05$ for CIGRE LV and $F(6, 14) = 2.09, p \geq 0.05$ for LV-rural3, homogeneity of variances was given. Next, we tested for normal distribution with the Shapiro-Wilk test, but this test's results were significant – $W = .7, p < .001$ for CIGRE LV and $W = .66, p < .001$ for LV-rural3 – and, thus, normality was not given. Since the sample size n was equal for all models, we relied on the robustness of ANOVA against violations of the assumption of normal distribution with equal sample sizes [(Ramachandran and Tsokos 2014) p. 512]. After that, we conducted a two-sided independent samples Welch's t-test on every possible pair of surrogate models. To reject the null hypothesis that the calculation times are identical, the tests are conducted towards a significance level of $\alpha = .05$. However, since we had more than two independent groups ($k = 7$), we applied the Bonferroni correction to prevent accumulation of alpha error.

Additionally, the measured calculation times were used to calculate a speed-up factor (SUF) in relation to the simulation model. The SUF measures how much lower the surrogate models' calculation time t_{sur} is compared to the simulation models' calculation time t_{sim} over all n simulation runs. Since the LV-rural3 grid was approximately three times as big as the CIGRE-LV grid model, the results could show whether the results generalize to grid models of different sizes. Due to the high number of variables involved, the comparison of the results was conducted in a qualitative manner.

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 7 of 12

Table 1 Results from the independent samples Welch's t-test for both grids. Shown are the t-statistics and the corresponding degree of freedom (*df*). Reported p-values were Bonferroni-corrected. Combinations missing in this table were significant with $p < .001$ for both grids
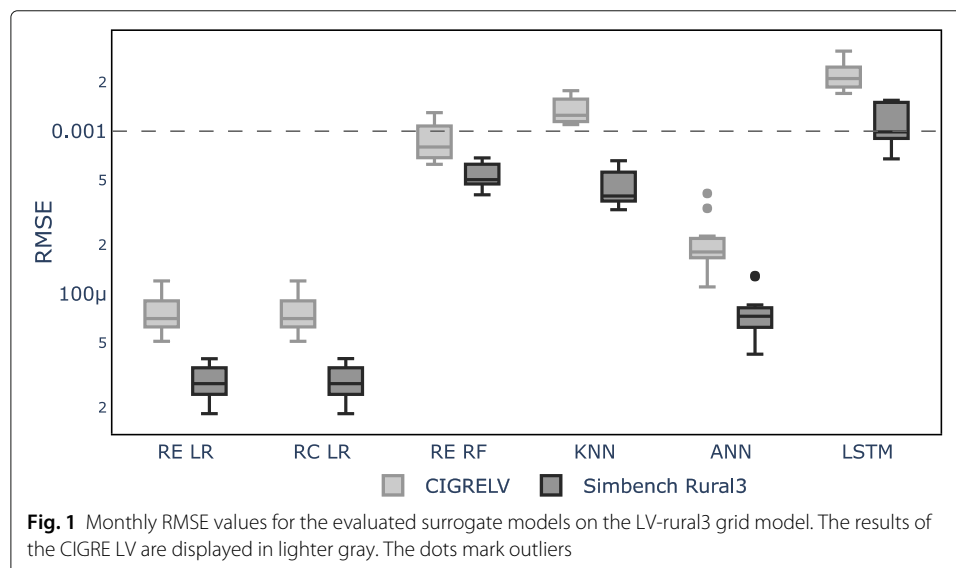
Model A	Model B	CIGRE LV		LV-rural3	
		df	*t*	*df*	*t*
Sim	RE LR	10.4	56.81***	2.0	25.54*
Sim	RE RF	16.7	-78.38***	2.1	-13.03
Sim	RC LR	11.2	60.83***	2.0	26.32*
Sim	k-NN	11.2	63.75***	2.0	26.02*
Sim	ANN	9.7	64.04***	2.0	27.77*
Sim	LSTM	10.4	1.38	2.0	27.58*
RE LR	RC LR	17.3	11.94***	3.8	17.25**
RE LR	k-NN	17.3	18.78***	2.2	13.02
RE LR	LSTM	18.0	-143.72***	2.6	53.87**
RC LR	k-NN	18.0	6.23***	2.2	-10.71
RC LR	ANN	14.5	2.22	3.9	40.01***
k-NN	ANN	14.5	-5.40**	2.3	72.84**
ANN	LSTM	16.4	-183.52***	3.2	-7.44

***$p < .001$
**$p < .01$
*$p < .05$

Results

Results of experiment 1

The results of the first experiment on the CIGRE LV, see Fig. 1, clearly showed large differences in the obtained error values between the different surrogate models. Especially the two models based on LR and the ANN achieved error values well below the defined criterion of an RMSE lower than 0.001, the LR even reaching RMSE values as low as 0.0001. Since the voltage magnitudes are given in the *pu*-system, this corresponds to error margins of 0.1% and 0.01%, respectively.

Fig. 1 Monthly RMSE values for the evaluated surrogate models on the LV-rural3 grid model. The results of the CIGRE LV are displayed in lighter gray. The dots mark outliers

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 8 of 12

When conducted on the second grid model, the surrogate models generally reached a lower RMSE value than they did on the CIGRE LV. Despite this apparent change in the results, the general order of surrogate models remained largely the same. The models based on LR still reached the best results, followed once again by the ANN. The RF and k-NN models reached fully satisfactory results this time, while the LSTM model still struggled to stay below the threshold.

Since overall results were better than they were on the smaller CIGRE LV, no negative impact on the prediction results by changes of the grid topology was apparent. Therefore, the results indicated that the use of surrogate models is robust towards changes in grid size.

We conclude the same statement towards the inclusion of a changing phase angle. While $cos(\varphi)$ was fixed at 0.9 in the CIGRELV, it was not constrained in the LV-rural3. However, this did not affect the surrogate models' ability to accurately map the simulation model's behavior. Additionally, the inclusion of volatile distributed energy generation in the form of PV generators did not have a negative effect on the surrogate models' accuracy.

All things considered, the experiment showed that results concerning accuracy obtained from CIGRE LV generalized remarkably well to grids with different factors. We could also show that there are large differences in the surrogate models' abilities to map the underlying grid simulation model. Especially the parametric models based on LR and the ANN performed well. The LSTM – despite being an extension of the neural network well suited to time series – did not manage to reach acceptable results on either grid model.

Results of experiment 2

The results obtained from the second experiment are meant to showcase the differences of calculation time between the surrogate models and the simulation model. In the ANOVA, the average calculation times over ten repetitions showed significant differences between the models ($F(6, 63) = 11291.78, p < .001$). Therefore, we conducted Welch's t-test, which is robust to a violation of normality (Hansen 2005), to determine pairwise differences. The results – see Table 1 – show significant differences between almost all model combinations. The only exceptions were marked by the pairs (Sim, LSTM) and (RC LR, ANN), which had no significant differences in the test. As with the first experiment, the ANOVA for LV-rural3 showed significant differences between the models concerning

Table 2 Overview of RMSE, mean calculation time (Calc. [s]), its standard deviation (StD.), and the models' speed-up factor (SUF) relative to the simulation model (Sim). Bold values mark the best of the respective column

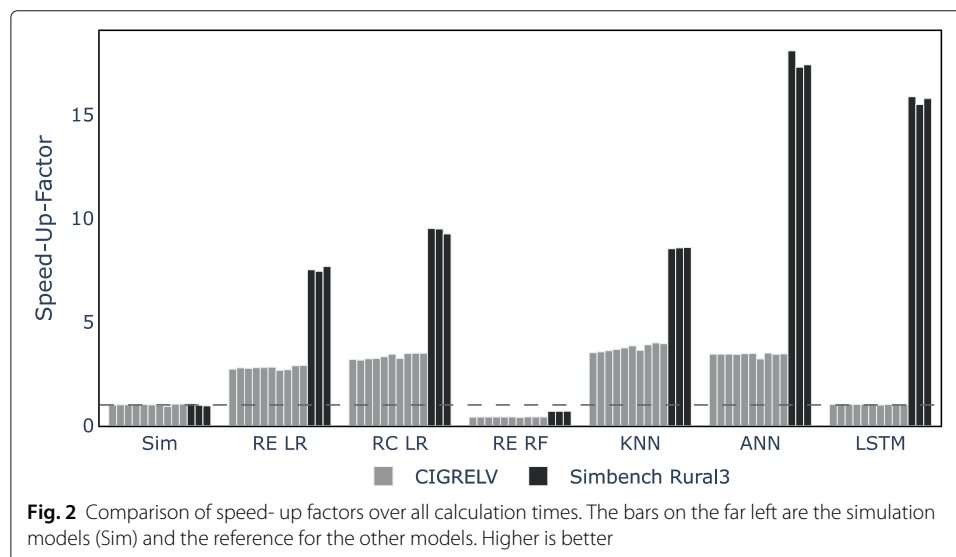|	CIGRELV				LV-rural3			
	RMSE	**Calc. [s]**	**StD.**	**SUF**	**RMSE**	**Calc. [s]**	**StD.**	**SUF**
Sim	-	1354.45	46.45	1.00	-	6388.01	375.63	1.00
RE LR	**7.57e-05**	486.64	13.21	2.78	**2.91e-05**	846.36	13.19	7.55
RC LR	**7.57e-05**	407.48	16.35	3.32	**2.91e-05**	678.38	10.50	9.42
RE RF	8.80e-04	3275.52	62.04	0.41	5.36e-04	9234.81	46.19	0.69
k-NN	1.34e-03	**362.26**	16.26	**3.74**	4.55e-04	745.28	2.58	8.57
ANN	2.08e-04	394.39	9.45	3.43	7.75e-05	**363.11**	8.71	**17.59**
LSTM	2.19e-03	1333.35	13.13	1.02	1.13e-03	406.46	5.10	15.72

calculation time and we could conduct Welch's *t*-test. The results in Table 1 showed significant differences between most of the model combinations, except for the pairs (Sim, RE RF), (RE LR, RC LR), (RC KR, k-NN), and (ANN, LSTM).

Table 2 illustrates that almost all models resulted in a considerable reduction of calculation time in comparison to the simulation model. The LSTM had almost the same calculation time as the simulation model, the RF model was unable to decrease the calculation time and was, in fact, slower than the simulation model. The fastest one of those was the k-NN model, which reached an average SUF of 3.74. It was closely followed by the ANN (3.43) and the LR models (3.32 and 2.78). For comparison, the reference model reached SUFs between 2.5 and 2.7 in a very similar experiment on the same grid model [(Balduin et al. 2019) p. 14].

Since the LV-rural3 model was far larger than the CIGRE LV, the simulation model showed a far longer calculation time, illustrated in Table 2. The same holds true for the majority of the surrogate models, which – except for the ANN and the LSTM – also took longer to calculate the relevant outputs. However, the increase in calculation time was less pronounced for the surrogate models. Therefore, all surrogate models reached higher SUFs when applied to the LV-rural3 than when applied to the smaller CIGRE LV. This effect was especially strong for the ANN and the LSTM, allowing them to reach SUFs of 17.59 and 15.72, respectively. With SUFs between 7.5 and 9.5, the remaining surrogate models achieved satisfactory results, as well. The results from both grid models, illustrated in Fig. 2, indicate that some models reacted more strongly to a larger grid model than others. Especially the ANN barely reacted to the larger size.

Conclusion

In this work, we evaluated different algorithms regarding their adequacy to be used as a surrogate model for a LV grid in two experiments. We changed some of the LV grid's parameters, namely the topology itself, the use of distributed energy generation, and the change from a constant to a varying phase angle. The changes were made to provide an estimation on the robustness regarding different grid topologies.

Fig. 2 Comparison of speed- up factors over all calculation times. The bars on the far left are the simulation models (Sim) and the reference for the other models. Higher is better

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 10 of 12

In the results of the first experiment we could verify that the LR-based models (RE LR and RC LR) and the ANN had a prediction error far below 0.1% on both used grid models, while RF on both grids and k-NN at least on the LV-rural3 grid still achieved satisfactory results. Only the LSTM did not pass the prediction error cutoff. Furthermore, the change of the topology and other parameters in the second grid model had nearly no effect on the general order of the surrogate models. From this, we deduced that the surrogate model algorithms were robust against parameter changes of the replaced simulation models.

The results of the second experiment showed the calculation time benefit of using surrogate models. Each surrogate model was significantly faster and provided a speed-up against the simulation model with one exception: the RF algorithm applied on the LV-rural3 was even slower than the simulation model. For the other models holds true that a change to a larger grid had actually increased the speed-up compared to the simulation model. whose calculation time was hardly influenced by the grid size. We conclude that a LR-based model or a k-NN model is good enough as a surrogate model for smaller grids, while ANN-based models further extended their advantage on larger grids. Considering the results obtained from both grid models, the experiment showed that a wide variety of different surrogate models can be used to decrease the calculation time of the simulation models.

Outlook

We addressed some of the open questions of the reference work, but some of them still are unanswered and new questions emerged. The tuning of hyperparameters could be further improved, e. g., penalizing the calculation time in the loss function, which could create a tendency towards smaller models. This would address the question of how much effort has to be made in order to obtain a useful surrogate model and how this effort relates to the benefits. Further improvements of the model could be considered, such as the inclusion of line loadings or other outputs of the grid simulation model. Another aspect would be to build a larger setup with a MV grid and several LV grids and investigate if the findings of this work still hold true.

Finally, the issue of limited training data should be addressed and we think that two approaches are conceivable. First, the available training data could be extended artificially by, e. g., the use of input distributions or bootstrapping. Secondly, the amount of training data required could be reduced by taking into account correlations of loads and generation in the surrogate modeling process. In our future work, we aim to address these question and, at the same time, include feedback from actual applications of our methodology in other related projects.

Authors' contributions
SB conceived and developed the methodology, implemented the simulation setup, and wrote most parts of the paper. TW developed the hypotheses, implemented the machine learning algorithms, conducted the experiments, and wrote

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 11 of 12

the corresponding sections of the paper. EP gave substantial feedback for the manuscript and contributed on the statistical analysis of the experiment results. All author(s) have read and approved the final manuscript.

Availability of data and materials
The datasets, results and code are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] OFFIS – Institute for Information Technology, Escherweg 2, 26121, Oldenburg DE, Germany.
[2] Helmut-Schmidt-Universität, Institut für Automatisierungstechnik, Holstenhofweg 85, 22043, Hamburg DE, Germany.

Published: 28 October 2020

References
Balduin S (2018) Surrogate models for composed simulation models in energy systems. Energy Inform 1(1):30

Balduin S, Oest F, Blank-Babazadeh M, Nieße A, Lehnhoff S (2019) Tool-assisted surrogate selection for simulation models in energy systems. In: 2019 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE. pp 185–192

Balduin S, Tröschel M, Lehnhoff S (2019) Towards domain-specific surrogate models for smart grid co-simulation. Energy Inform 2(1):27

Balduin S, Westermann T, Puiutta E (2020) Evaluating different machine learning techniques as surrogate for low voltage grids. arXiv preprint arXiv:2006.12389

Baumann M, Grundel S, Sauerteig P, Worthmann K (2019) Surrogate models in bidirectional optimization of coupled microgrids. at-Automatisierungstechnik 67(12):1035–1046

Beisheim B, Rahimi-Adli K, Krämer S, Engell S (2019) Energy performance analysis of continuous processes using surrogate models. Energy 183:776–787

Blank M, Gandor M, Niesse A, Scherfke S, Lehnhoff S, Sonnenschein M (2015) Regionally-specific scenarios for smart grid simulations. In: 5th International Conference on Power Engineering, Energy and Electrical Drives (POWERENG2015). IEEE. pp 250–256. http://dx.doi.org/10.1109/PowerEng.2015.7266328

Blank M, Lehnhoff S (2014) Correlations in reliability assessment of agent-based ancillary-service coalitions. In: 2014 Power Systems Computation Conference. IEEE. pp 1–7

Danner D, de Meer H (2019) State estimation in the power distribution system. ACM SIGMETRICS Perform Eval Rev 46(3):86–88

Dimitrov N (2019) Surrogate models for parameterized representation of wake-induced loads in wind farms. Wind Energy 22(10):1371–1389

Donon B, Clément R, Donnot B, Marot A, Guyon I, Schoenauer M (2020) Neural networks for power flow: Graph neural solver. Electr Power Syst Res 189:106547

Fischer L, Memmen J-M, Veith EM, Tröschel M (2019) Adversarial resilience learning – towards systemic vulnerability analysis for large and complex systems. Ninth Int Conf Smart Grids Green Commun IT Energy-aware Technol 9(1):24–32

Frank S, Steponavice I, Rebennack S (2012) Optimal power flow: a bibliographic survey ii. Energy Syst 3(3):259–289

Gerster J (2018) Agent-based dynamic optimization of local controller configurations in converter dominated electricity grids using decoder functions. Energy Inform 1(1):46

Grisales-Noreña L, Garzon-Rivera O, Ramırez-Vanegas C, Montoya O, Ramos-Paja C (2020) Application of the backward/forward sweep method for solving the power flow problem in dc networks with radial structure. In: Journal of Physics: Conference Series, vol. 1448. p 012012

Grundel S, Sauerteig P, Worthmann K (2019) Surrogate models for coupled microgrids. In: Progress in Industrial Mathematics at ECMI 2018. Springer, Cham, Switzerland. pp 477–483

Gupta S, Kazi F, Wagh S, Kambli R (2015) Neural network based early warning system for an emerging blackout in smart grid power networks. In: Buyya R, Thampi SM (eds). Intelligent Distributed Computing. Springer, Cham. pp 173–183

Hansen J (2005) Using SPSS for Windows and Macintosh: Analyzing and Understanding Data (4th ed.) Am Stat 59(1):113

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Kontis EO, Kryonidis GC, Nousdilis AI, Malamaki K-ND, Papagiannis GK (2019) Power flow analysis of islanded ac microgrids. In: 2019 IEEE Milan PowerTech. IEEE. pp 1–6

Montoya O, Escobar A, Garrido V (2020) Power flow solution in direct current grids using the linear conjugate gradient approach. In: Journal of Physics: Conference Series, vol. 1448. p 012016

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

Nagpal S, Mueller C, Aijazi A, Reinhart CF (2019) A methodology for auto-calibrating urban building energy models using surrogate modeling techniques. J Build Perform Simul 12(1):1–16

Nilsson M, Soder L, Olauson J, Eriksson R, Nordström L, Ericsson GN (2018) A machine learning method creating network models based on measurements. In: 2018 Power Systems Computation Conference (PSCC). IEEE. pp 1–7

Papathanassiou S, Hatziargyriou N, Strunz K, et al (2005) A benchmark low voltage microgrid network. In: Proceedings of the CIGRE Symposium: Power Systems with Dispersed Generation. CIGRE. pp 1–8

Balduin *et al. Energy Informatics* 2020, **3**(Suppl 1):24

Page 12 of 12

Puiutta E, Veith EM (2020) Explainable reinforcement learning: A survey. arXiv preprint arXiv:2005.06247

Ramachandran KM, Tsokos CP (2014) Mathematical statistics with applications in r. Elsevier, San Diego, CA, USA. Elsevier

Simpson TW, Poplinski JD, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: Survey and recommendations. Engineering with Computers 17(2):129–150

Slot RM, Sørensen JD, Sudret B, Svenningsen L, Thøgersen ML (2020) Surrogate model uncertainty in wind turbine reliability assessment. Renew Energy 151:1150–1162

Spalthoff C, Sarajlic D, Kittl C, Drauz S, Kneiske T, Rehtanz C, Braun M (2019) Simbench: Open source time series of power load, storage and generation for the simulation of electrical distribution grids. In: International ETG-Congress 2019; ETG Symposium. VDE. pp 1–6

Spyromitros-Xioufis E, Tsoumakas G, Groves W, Vlahavas I (2016) Multi-target regression via input space expansion: treating targets as inputs. Mach Learn 104(1):55–98

Steinbrink C (2016) A nonintrusive uncertainty quantification system for modular smart grid co-simulation. PhD thesis, Ph. D. thesis, University of Oldenburg

Steinbrink C, Lehnhoff S, Rohjans S, Strasser TI, Widl E, 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. Springer. pp 171–185

Syai'in M, Soeprijanto A (2010) Neural network optimal power flow (nn-opf) based on ipso with developed load cluster method. World Acad Sci Eng Technol 72:48–53

Thurner L, Scheidler A, Schafer F, Menke JH, Dollichon J, Meier F, Meinecke S, Braun M (2018) pandapower - an open source python tool for convenient modeling, analysis and optimization of electric power systems. IEEE Trans Power Syst 33(6):6510–6521

Vazquez-Canteli J, Demir AD, Brown J, Nagy Z (2019) Deep neural networks as surrogate models for urban energy simulations. In: Journal of Physics: Conference Series, vol. 1343. IOP Publishing. p 012002

Veerasamy V, Wahab NIA, Ramachandran R, Madasamy B, Mansoor M, Othman ML, Hizam H (2020) A novel rk4-hopfield neural network for power flow analysis of power system. Appl Soft Comput:106346

Veith EM, Wenninghoff N, Frost E (2020) The adversarial resilience learning architecture for ai-based modelling, exploration, and operation of complex cyber-physical systems. http://arxiv.org/abs/2005.13601. Accessed 1 Oct 2020

Xiang M, Yu J, Yang Z, Yang Y, Yu H, He H (2020) Probabilistic power flow with topology changes based on deep neural network. Int J Electr Power Energy Syst 117:105650

Yuan Z, Wang Y, Yi Y, Wang C, Zhao Y, Wen W (2019) Fast linear power flow algorithm for the study of steady-state performance of dc grid. IEEE Trans Power Syst 34(6):4240–4248

Zhao J, Gómez-Expósito A, Netto M, Mili L, Abur A, Terzija V, Kamwa I, Pal B, Singh AK, Qi J, et al (2019) Power system dynamic state estimation: Motivations, definitions, methodologies, and future work. IEEE Trans Power Syst 34(4):3188–3198

Publisher's Note

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