EVALUATING DIFFERENT MACHINE LEARNING TECHNIQUES AS SURROGATE FOR LOW VOLTAGE GRIDS

Stephan Balduin  
OFFIS – Institute for Information Technology  
Escherweg 2, 26121 Oldenburg  
stephan.balduin@offis.de

Tom Westermann  
OFFIS – Institute for Information Technology  
tom.westermann@outlook.de

Erika Puiutta  
OFFIS – Institute for Information Technology  
Escherweg 2, 26121 Oldenburg  
erika.puiutta@offis.de

June 19, 2020

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, a surrogate model for a low voltage grid was built 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

1 Introduction

The ongoing transformation of 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, the possibilities to test new technologies are very restricted and, since it is also an expensive infrastructure, the installation of a large-scale test grid is not feasible. For this reason, simulation and hardware-in-the-loop are used for the development and testing of such new technologies 1. The so-called smart grid comprises not only the power grid, but also other domains, such as the Information and Communication Technology (ICT) domain and the gas and heat energy systems. Usually, for each domain, experts are developing their own simulation environment. One smart solution to couple these different domains incorporates the use of co-simulation, which only requires to implement interfaces for each domain 2. In such a setup, synchronization and data exchange between different environments is handled by a co-simulation framework (e.g., the co-simulation framework mosaik1).

Even with co-simulation, building such a large, cross-domain simulation environment is still a complex task. In addition, depending on the size and complexity of the underlying scenario, the simulation of the overall system can become very

1https://mosaik.offis.de retrieved on 10 Jun. 2020
slow [3]. 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 [4].

In a recent work, a surrogate model for a subsystem of such a multi-domain co-simulation environment was built [5]. Starting with the Conseil International des Grands Réseaux Electriques (CIGRE low voltage (LV) benchmark power grid model [6], the authors added several load models that were connected to certain buses in the grid. A surrogate model was built to replace the load models and the power grid model. The authors evaluated the surrogate model in a case study with several experiments. Their 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.

Although we think this is a promising approach, the authors concluded their work with several simplifying assumptions and open questions, from which we recap the most important ones. First, the simulation setup consisted only of load models, but in a smart grid scenario, there would also be distributed energy resources (DER) like photovoltaic (PV) or combined heat and power (CHP). Secondly, the authors described the architecture of their deep neural network (DNN) surrogate model as simple, which we will discuss later in this paper. Furthermore, other machine learning (ML) techniques could be better suited to this particular problem. Finally, although the proposed methodology worked for the CIGRE LV grid, it would be interesting to know if this methodology works for other grid models, as well.

In this paper, we aim to address these issues. 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 that served as the basis for our study. In addition, we discuss the modifications we made to this simulation setup. 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 section Conclusion. Finally, we discuss open questions and future work in the section Outlook.

2 Related Work

Surrogate models – in the literature also referred to as metamodels, response surfaces, or emulators – are a common technique from the field of statistical design of experiments [7]. They are used to describe the behavior of a system that, for various reasons, is not suited to be built knowledge-based. Surrogate models can be found in many domains, but we will focus on the energy domain. They are used in a broad range of use cases: starting from the calculation and optimization of energy savings [8, 9, 10] and the replacement of specific simulation models [11, 12] over surrogate models for (micro)grids [13, 14, 15] to the use in uncertainty and reliability assessment [16, 17, 18]. This list is far from complete and there are also other approaches such as in Gerster [18] 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. PF analysis is performed in different use cases; market analysis and short-term operational planning are only two of them. The PF analysis part used in this paper assumes that the power grid is in a steady-state and has no special use case in mind despite the PF analysis itself. 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.

Grisales-Noreña et al. [19] compared six different methods from the literature with their approach of backward/forward sweep iteration for the PF calculation in direct current (DC) grids with radial structure. The authors benchmarked their approach on four different grid sizes and achieved a performance increase, especially on the larger grids. Another approach from Montoya et al. [20] aimed to overcome the issue of the costly matrix inversions that are normally required to solve the PF. Montoya et al. proposed a classical gradient conjugate method to solve linear algebraic equations in DC grids. Kontis et al. [21] investigated the PF in islanded DC microgrids that are operated under the droop control scheme. A steady-state analysis is not easy to employ for islanded grids since there is no slack bus, but the proposed method showed promising results regarding accuracy and robustness. Yuan et al. [22] proposed a linear redefinition of the nonlinear power flow equations for DC grids. Although not applicable to alternating current (AC) grids, the authors reported a calculation speed twenty times faster than using NR on grids with more than a hundred nodes and an extremely small error.

But there is also a broad variety of approaches to solve the PF and related problems with data-driven algorithms. Nilsson et al. [23] 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.
The work in [24] created a modified Hopfield artificial neural network (ANN) to solve PF equations. The results were evaluated on different standard IEEE power systems. Another example is given by [25] who proposed an ANN based on a Graph Neural Solver. Frank et al. [26] surveyed about data-driven approaches to the closely related optimal power flow (OPF) problem. In [27], the authors used a DNN to overcome the high computational burden of OPF. Syai'in and Soeprijanto [28] proposed their method called ANN OPF, compared it to improved particle swarm optimization, and achieved a faster response time. Gupta et al. [29] used a neural network to predict cascading failures in advance.

One question when applying ML to power grid models is related to the gathering of training data. In [5], the authors pointed out the challenges of the sampling process for power grids. They tried to replace their set of load profiles using an empirical sampling strategy with kernel density estimation, but this method did not provide satisfactory results. Therefore, the authors did not fully replace their initial data but extended them with sampled data. Danner and de Meer [30] 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. Danner and de Meer used the sampled data to build a dependency graph, which is used as input for different ML techniques. 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. [31]. They compared techniques from the fields of model reduction and machine learning, which are quite similar but with subtle differences, to build parametric surrogate models. The two case studies presented by Zhao et al. considered the prediction of pressure fields around an airfoil and the prediction of strain field over a damaged composite panel. 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. Though their work is not located in the energy domain, the approach could also be useful to mitigate the sampling problem for power grids. 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 [32].

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) [33] utilizes two classes of RL agents competing in a shared environment. In the simplest case, one agent tries to break the environment, and the other one aims to keep the environment in a healthy state [34]. The ARL methodology could be used to explore the sampling space, localize limitations of the model, and identify critical spots to derive sampling rules.

### 3 Simulation Setup

The simulation setup in [5] used the pandapower [35] implementation of the CIGRE LV benchmark grid [6]. This grid consisted of three subgrids: a residential subgrid, an industrial subgrid, and a commercial subgrid. A time series was assigned to each of the load models. In each step, these models forwarded the corresponding value from the time series to the grid. For the residential area, a data set from the research project Smart Nord [36] was used that consisted of synthesized time series of households. In total, these time series approximated the default load profile that is commonly used in Germany to approximate household consumer load. For the industrial and commercial areas, a reference data set from openei [37] was used, which consisted of time series of several commercial consumers.

The load models and the pandapower grid model were coupled using the co-simulation framework mosaik. While in the industrial and commercial subgrids to each load of the grid one load model was connected to, in the residential subgrid several load models could be connected to the same load node. This assignment was considered as domain knowledge about the grid as well as other topological information such as which load node is connected to which bus. The surrogate model has not received this information, but only time series as input, i.e., in this regard, the model was built without domain knowledge. Figure 1 illustrates this relationship and shows what is and what is not part of the surrogate model. This decision – to distinguish between time series and a value-forwarding load model – was made with the goal to incorporate PVs and CHPs as part of this system to-be-replaced as well. Using unassigned load data and, in the future, the weather forecast for DER, the surrogate model should predict the voltage magnitude per unit (vm_pu) of the grid’s buses. In a larger setup, this would add the value of a better estimation and more flexibility of this LV grid regarding distributed energy generation in comparison to a setup where this grid is replaced by a single time series.

The surrogate model was built using a DNN, which consists of several fully connected (dense) hidden layers. A random-search cross-validation approach was used to evaluate a small set of hyperparameters. Among others, this included the number of hidden layers. However, the number of neurons in each hidden layer was determined depending

---

[24] Balduin et al. (2020)

[25] Frank et al. (2020)

[27] Syai'in and Soeprijanto (2020)

[29] Gupta et al. (2020)

[30] Danner and de Meer (2020)

[31] Zhao et al. (2020)

[32] See [32] for a deeper insight in the field of explainability of AI and RL.

[33] ARL methodology could be used to explore the sampling space, localize limitations of the model, and identify critical spots to derive sampling rules.

[35] Pandapower implementation of the CIGRE LV benchmark grid.

[36] Smart Nord project.

[37] OpenEI dataset.

---

[https://openei.org/datasets/files/961/pub](https://openei.org/datasets/files/961/pub) retrieved on 10 Jun. 2020
Figure 1: Simulation setup used in the referenced work is represented by the left part. Data from time series (orange circles) were transmitted to the load models (blue circles), from the load models to one of the load aggregators (grey circle with the \(\sum\) symbol), and from there to the grid model (large grey circle). The grid model performed PF calculations and the results were gathered, again, as a time series. In the referenced work, the models inside the bold black box were replaced by a surrogate model. For this work, we extended the scenario for the right part and built surrogate models for the models inside the dashed black box.

on the sizes of input- and output-layer as well as the number of hidden layers, based on a rule-of-thumb [37]. The resulting model was tested in different experiments evaluating the accuracy, a speed-up factor (SUF), and the capability to correctly predict voltage violations at the grid’s buses. Since the entire architecture was created generically with few, domain-independent parameters, the authors concluded that the model architecture was simple and should be further improved.

For this work, we aimed to put some domain knowledge into the DNN and compared the new model to the reference model from the previous work. Furthermore, we evaluated additional ML models to find the most suitable model for this simulation setup. Therefore, we included ML models with different characteristics ranging from single target predictor ensembles to multi-target models and recurrent neural networks (RNN), resulting in a total of six different surrogate models. We discuss these models further in section [Model selection]. 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 [38] 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. Using this simbench grid, we aimed to address the open questions of the previous work.

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. Figure 2 shows the aggregation of the residential time series on the left and the commercial time series on the right. 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, which, unlike the residential loads, tend to be higher in summer. The data of the simbench project are shown in Figure 3. On the left are, again, the residential loads. These show a behavior similar to the Smart Nord loads, with the tendency to be higher in winter. Instead of commercial loads, the simbench project provides time series for PV generation, which are shown on the right. The PV generation is higher in summer, remarkably low in December and January, and heavily depending on the weather conditions.

Finally, we compared the different topologies, which are illustrated in Figure 4 and Figure 5. 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.

4 Case Study

Based on this setup, we conducted a case study to validate our extensions against the open questions of the previous work. We will discuss these questions in the first part of this section and then hypothesize to define the context of our experiments. The second part of this section consists of a brief overview of the ML models, their parameter tuning strategies, and the reasons for their selection. Afterwards, we describe the data generation and the experimental setup in detail. Finally, this section concludes with a description of the experiments’ results.

4.1 Hypothesis

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 [5].

To compare these new models to the reference model, we defined three hypotheses similar to the reference hypothesis. The first one deals with the prediction quality of the models. To be predicted were the \( \text{vm}_\text{pu} \) of the buses, which were normalized values that should lie in the interval \([0.9, 1.1]\). In the reference work, a quality threshold of \(10\%\) was used but the authors concluded that the model was not accurate enough. Therefore, we will set the threshold to \(0.1\%\).
Figure 4: Topology of the CIGRE LV benchmark grid, from [5]. The blue boxes represent load nodes, with different load types indicated by the symbol inside the box. Lines represent lines of the grid, except for the bold lines, which represent bus bars.

The reasoning behind this decision as well as the used metric for the error are discussed in the section Description of Experiment 1.

**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 $\leq 0.1\%$

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. Since we did not know in advance if the reference model was exemplarily good or bad regarding the computation time and compared to other models, we compared all surrogate models with the original simulation model. 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. Additionally, several simplifications were made for the reference environment. Only load models were used and, though with different characteristics, a constant phase angle was used. We selected the simbench LV-rural3 data set since it provides PV power generation and a varying phase angle as well as a totally different grid topology. We discuss the difference in more detail in the section Data generation. 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. We provide more information on this topic in section Description of Experiment 2.

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, since there are so many variables involved.
- instead, a qualitative comparison between two simulation models will be performed.

4.2 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 is a well established and widely known algorithm to build a regression model. LR requires a fairly low number of calculation steps and, since only one output is calculated, the result is interpretable and explainable. Another single-target model is the random forest (RF), which is an ensemble method of decision trees. The RF algorithm is more flexible than regular LR due to its ability to also model non-linear input-output relationships. However, large numbers of trees slow down the training process. In contrast to regular LR, RFs have several hyperparameters. The three most important ones are candidate variables per split (mtry), the sample size of each tree’s training process, and the minimum number of observations in an endnode (nodesize) [40]. 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 [41]. 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), which is rather simple but often provides good results. This algorithm 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. K-NN provides a fast training process and good prediction speed. The second multi-target model we selected was the long short-term memory (LSTM) network, which is an adaptation of ANNs specifically suited to temporal data [42]. 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 the 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.

4.3 Experimental setup

The following section deals with the setup of the experiments. We start with a description of the generation of training data. The section will be followed up by a description of two experiments we conducted to test the three hypotheses that were established earlier.

4.3.1 Data generation

To train surrogate models, a sufficiently large training and testing data set was needed. Since both the simulation models and matching realistic load time series were available, calculating the \( \text{vm}_{\text{pu}} \) of the bus bars was a straightforward task. Instead of implementing an elaborated experimental design that seeks to map the entire input space, the input values were restricted to realistic combinations of input values. This marks a deviation from the sampling process used in the reference work. The input and output values were obtained within the simulation setup described in the section Simulation Setup, whereby all loads (active and reactive power) and all \( \text{vm}_{\text{pu}} \) were logged at every step. In this way, the entire span of the time series (365 days) was used to calculate a set of input and output values, which could be used to train the models.

There were a number of differences between the two simulation models that had an effect on the process of data generation. Since one of the assumptions for the CIGRE LV was a constant phase angle \( \cos(\varphi) = 0.9 \) between the voltage and the current in the grid, the reactive power \( Q \) for the input values was calculated from the active power \( P \). The LV-rural3, on the other hand, had separate time series for active and reactive power consumption, thus it featured a variable phase angle \( \cos(\varphi) \). Additionally, the LV-rural3 included numerous distributed energy generators in the form of PV systems, each of which had its own time series for active and reactive power fed into the grid. These time series were also used as input values for the simulation models.

4.3.2 Description of Experiment 1

The goal of the first experiment was to verify whether the surrogate models were able to accurately predict the \( \text{vm}_{\text{pu}} \) calculated by the simulation model. Since we intended to explore the maximum accuracy potential of the machine learning algorithms, a training set of maximum size was used. However, since one of the surrogate models was based on an LSTM network, randomly splitting the data into training and testing set eliminates temporal information required by this 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. Thus, we could evaluate the models over the span of the entire year. Due to the lack of a larger data set and the cyclical nature of the data, we used observations from the months following the test month as well. To quantify the prediction error, the average root mean squared error (RMSE) over all \( n \) observations and all \( m \) bus \( \text{vm}_{\text{pu}} \) of the LV grid was calculated.

\[
RMSE = \sqrt{\frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} - \hat{y}_{ij})^2}
\]  

At this, \( y_{ij} \) is the actual value and \( \hat{y}_{ij} \) the prediction of the \( \text{vm}_{\text{pu}} \) of bus \( j \) in observation \( i \). The RMSE was used as a criterion to decide whether the null hypothesis can be rejected. The more challenging aspect of this criterion was to determine an adequate threshold value. Different applications of the surrogate models could potentially have completely different requirements in terms of accuracy and reliability. Furthermore, too high accuracy on the training
set may be the consequence of overfitting. 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 $10^{-3}$ ($\text{RMSE}_{\text{year}} \leq 10^{-3}$). 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. We then quantitatively compared and examined the experiments’ results to find possible effects caused by the changed parameters.

### 4.3.3 Description of Experiment 2

The second experiment aimed to provide insight into the surrogate models’ calculation speed when used in a co-simulation-setup. In order to provide a baseline calculation time, the same experiment was conducted with the simulation models. 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. Since the only difference between the setups is the choice of the surrogate model, all substantial changes in calculation time could be traced back to them. 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. However, 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 [43, p. 512]. After that, we conducted a two-sided independent samples Welch’s t-test on every possible pair of surrogate models. With the results of these tests it was possible to determine whether the average calculation time differs between two 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 SUF in relation to the simulation model. The SUF measures how much lower the surrogate models’ calculation time $t_{\text{sur}}$ is compared to the simulation models’ calculation time $t_{\text{sim}}$ over all $n$ simulation runs.

$$\text{SUF} = \frac{\frac{1}{n} \sum_{i=1}^{n} t_{\text{sim},i}}{\frac{1}{n} \sum_{j=1}^{n} t_{\text{sur},j}}$$

(2)

As it was the case with the first experiment, we repeated the second one on the LV-rural3 grid model. Since it 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.

### 4.4 Results

#### 4.4.1 Results of Experiment 1

The results of the first experiment on the CIGRE LV, see Figure 6, 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. The RF model reached an RMSE value just below the threshold of 0.001, while the k-NN model reached a value just above it. The LSTM, on the other hand, was not able to achieve a sufficiently low error value on the full grid.

A closer evaluation of the RMSE values showed that the surrogate models’ performance varied strongly between the different subgrids of the grid model. In the commercial and industrial subgrids, most models achieved lower error rates
than in the residential subgrid. Due to the weaker results in the residential subgrid, the only surrogate models that reach satisfactory results in every subgrid were the surrogate models based on LR and the ANN. However, both of them stayed well below the defined threshold.

When conducted on the second grid model – the LV-rural3, see the results in Figure 7 – 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.

Figure 8 shows the differences between the surrogate models’ results. Especially in the rapidly changing \( \text{vm}_{\text{p}} \text{s} \) in the CIGRE LV, the less accurate models struggled to accurately map the simulation models’ behavior. The models based on LR and the ANN, on the other hand, barely showed any visible discrepancies.

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. Feeding the generated energy into the surrogate model in the same way as the loads appeared to be a viable solution to the inclusion of energy generated in a decentralized manner.

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
Figure 7: 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 for comparison. The dots mark outliers.

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. A possible explanation would be the large 15-minute intervals between measurements. Due to the more rapid fluctuations in the voltage magnitudes and loads, the greater flexibility of the LSTM seemed to be more of an obstacle than a help due to the limited amount of relevant information in the preceding time steps.

4.4.2 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 [44], 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 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 [5, 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 [Figure 9] indicate that
Figure 8: Comparison of simulated \textit{vm\_p}us and surrogate models’ predictions. The left column shows a results for one day in the residential subgrid of the CIGRE LV, while the right column shows the same for a bus in the LV-rural3. Both buses obtained the highest error values in their respective grid models.

some models reacted more strongly to a larger grid model than others. Especially the ANN barely reacted to the larger size.

5 Conclusion

In this work, we evaluated different algorithms regarding their adequacy to be used as a surrogate model for a LV grid. These algorithms can be categorized into ensembles of single-target models (LR and RF) and multi-target models (k-NN, ANN, and LSTM), models with low (k-NN and RF) to high (ANN and LSTM) number of hyperparameters, distance-based models (k-NN), or neural network models (ANN and LSTM). All surrogate models were evaluated in two experiments, the first one regarding the accuracy and the second one regarding the calculation time compared to the simulation model. Additionally, 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. Therefore, we conducted the two experiments on two different grid models: the CIGRE LV benchmark grid and the simbench LV-rural3 grid.

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 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. On the CIGRE LV grid, the prediction error of the models differed significantly depending on the subgrid. We attribute this to more regular behavior of the commercial data sets. 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. This could be caused by a number of trees that was chosen too high. For the other models holds true that a change to a larger grid had actually increased the speed-up compared to the simulation model. This applied in particular to the ANN-based models 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.

From both experiments, we concluded that the use of an ANN as the surrogate model in the reference work was not a bad choice although the evidence was not provided in that very work. Additionally, we improved the model.

---

3The higher calculation time of the LSTM network on the CIGRE LV can likely be attributed to the use of a second (calculation intensive) recurrent layer that was a result from hyperparameter tuning.
architecture to provide an even lower error and a higher speed-up. Furthermore, we could verify that a change of the grid topology and other parameters like the integration of distributed energy resources and a varying phase angle had no negative impact on the quality of the model.

6 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. In the reference work, there was an experiment regarding the detection of critical voltage violations, which could not be transferred to LV-rural3 since there were no voltage violations inside the data. The problem could be circumvented by adding more data to the grid and then conducting the corresponding experiments again. 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.

Acknowledgements

Parts of this work have been developed in the context of a master thesis at the Carl v. Ossietzky University of Oldenburg and the Helmut-Schmidt University of Hamburg. We thank our supervisors Prof. Dr.-Ing Alexander Fay, Prof. Dr. Sebastian Lehnhoff and Jan-Philip Beck. We also thank our colleagues Johannes Gerster, Lasse Hammer, Daniel Lange and Eric Veith for their helpful comments and contributions during the preparation of this manuscript.
References

[1] Cornelius Steinbrink, Sebastian Lehnhoff, Sebastian Rohjans, Thomas I Strasser, Edmund Widl, Cyndi Moyo, Georg Lauss, Felix Lehfuß, Mario Faschung, Peter Palensky, et al. Simulation-based validation of smart grids–status quo and future research trends. In International Conference on Industrial Applications of Holonic and Multi-Agent Systems, pages 171–185. Springer, 2017.

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

[3] Marita Blank, Malin Gandor, Astrid Niesse, Stefan Scherfke, Sebastian Lehnhoff, and Michael Sonnenschein. Regionally-specific scenarios for smart grid simulations. In 5th International Conference on Power Engineering, Energy and Electrical Drives (POWERENG2015), pages 250–256. IEEE, 5 2015.

[4] T.W. Simpson, J.D. Poplinski, P. N. Koch, and J.K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. Engineering with Computers, 17(2):129–150, 2001.

[5] Stephan Balduin, Martin Tröschel, and Sebastian Lehnhoff. Towards domain-specific surrogate models for smart grid co-simulation. Energy Informatics, 2(1):27, 2019.

[6] Stavros Papathanassiou, Nikos Hatziargyriou, Kai Strunz, et al. A benchmark low voltage microgrid network. In Proceedings of the CIGRE symposium: power systems with dispersed generation, pages 1–8. CIGRE, 2005.

[7] Raymond H Myers, Douglas C Montgomery, and Christine M Anderson-Cook. Response surface methodology: process and product optimization using designed experiments. John Wiley & Sons, 2016.

[8] Benedikt Beisheim, Keivan Rahimi-Adli, Stefan Krämer, and Sebastian Engell. Energy performance analysis of continuous processes using surrogate models. Energy, 183:776–787, 2019.

[9] Jose Vazquez-Canteli, Aysegul Dilsiz Demir, Julien Brown, and Zoltan Nagy. Deep neural networks as surrogate models for urban energy simulations. In Journal of Physics: Conference Series, volume 1343, page 012002. IOP Publishing, 2019.

[10] Shreshth Nagpal, Caitlin Mueller, Arfa Aijazi, and Christoph F Reinhart. A methodology for auto-calibrating urban building energy models using surrogate modeling techniques. Journal of Building Performance Simulation, 12(1):1–16, 2019.

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

[12] Stephan Balduin, Frauke Oest, Marita Blank-Babazadeh, Astrid Nieße, and Sebastian Lehnhoff. Tool-assisted surrogate selection for simulation models in energy systems. In 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), pages 185–192. IEEE, 2019.

[13] Sara Grundel, Philipp Sauerteig, and Karl Worthmann. Surrogate models for coupled microgrids. In Progress in Industrial Mathematics at ECM 2018, pages 477–483. Springer, 2019.

[14] Manuel Baumann, Sara Grundel, Philipp Sauerteig, and Karl Worthmann. Surrogate models in bidirectional optimization of coupled microgrids. at-Automatisierungstechnik, 67(12):1035–1046, 2019.

[15] Stephan Balduin. Surrogate models for composed simulation models in energy systems. Energy Informatics, 1(1):30, 2018.

[16] René MM Slot, John D Sørensen, Bruno Sudret, Lasse Svenningsen, and Morten L Thøgersen. Surrogate model uncertainty in wind turbine reliability assessment. Renewable Energy, 151:1150–1162, 2020.

[17] Marita Blank and Sebastian Lehnhoff. Correlations in reliability assessment of agent-based ancillary-service coalitions. In 2014 Power Systems Computation Conference, pages 1–7. IEEE, 2014.

[18] Johannes Gerster. Agent-based dynamic optimization of local controller configurations in converter dominated electricity grids using decoder functions. Energy Informatics, 1(1):46, 2018.

[19] LF Grisales-Noreña, OD Garzon-Rivera, CA Ramirez-Vanegas, OD Montoya, and CA Ramos-Paja. 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, volume 1448, page 012012, 2020.

[20] OD Montoya, AF Escobar, and VM Garrido. Power flow solution in direct current grids using the linear conjugate gradient approach. In Journal of Physics: Conference Series, volume 1448, page 012016, 2020.

[21] Elefherios O Kontis, Georgios C Kryonidis, Angelos I Nousdilis, Kyriaki-Nefeli D Malamaki, and Grigoris K Papagiannis. Power flow analysis of islanded ac microgrids. In 2019 IEEE Milan PowerTech, pages 1–6. IEEE, 2019.
[22] Zhichang Yuan, Yizhen Wang, Yue Yi, Chengshan Wang, Yuming Zhao, and Weijie Wen. Fast linear power flow algorithm for the study of steady-state performance of dc grid. *IEEE Transactions on Power Systems*, 34(6):4240–4248, 2019.

[23] Martin Nilsson, Lennart Soder, Jon Olauson, Robert Eriksson, Lars Nordström, and Göran N Ericsson. A machine learning method creating network models based on measurements. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7. IEEE, 2018.

[24] Veerapandiyan Veerasamy, Noor Izzi Abdul Wahab, Rajeswari Ramachandran, Balasubramonian Madasamy, Muhammad Mansoor, Mohammad Lutfi Othman, and Hashim Hizam. A novel rk4-hopfield neural network for power flow analysis of power system. *Applied Soft Computing*, page 106346, 2020.

[25] Balthazar Donon, Rémy Clément, Benjamin Donnot, Antoine Marot, Isabelle Guyon, and Marc Schoenauer. Neural networks for power flow: Graph neural solver. 2019.

[26] Stephen Frank, Ingrida Steponavice, and Steffen Rebennack. Optimal power flow: a bibliographic survey ii. *Energy Systems*, 3(3):259–289, 2012.

[27] Mingxu Xiang, Juan Yu, Zhifang Yang, Yan Yang, Hongxin Yu, and He He. Probabilistic power flow with topology changes based on deep neural network. *International Journal of Electrical Power & Energy Systems*, 117:105650, 2020.

[28] Mat Syai’in and Adi Soepriyanto. Neural network optimal power flow (nn-opf) based on ipso with developed load cluster method. *World Academy of Science, Engineering and Technology*, 72:48–53, 2010.

[29] Sudha Gupta, Faruk Kazi, Sushama Wagh, and Ruta Kambli. Neural network based early warning system for an emerging blackout in smart grid power networks. In Rajkumar Buyya and Sabu M. Thampi, editors, *Intelligent Distributed Computing*, pages 173–183, Cham, 2015. Springer International Publishing.

[30] Dominik Danner and Hermann de Meer. State estimation in the power distribution system. *ACM SIGMETRICS Performance Evaluation Review*, 46(3):86–88, 2019.

[31] Junbo Zhao, Antonio Gómez-Expósito, Marcos Netto, Lamite Mili, Ali Abur, Vladimir Terzija, Innocent Kamwa, Bikash Pal, Abhinav Kumar Singh, Junjian Qi, et al. Power system dynamic state estimation: Motivations, definitions, methodologies, and future work. *IEEE Transactions on Power Systems*, 34(4):3188–3198, 2019.

[32] Erica Puiutta and Eric MSP Veith. Explainable reinforcement learning: A survey. *arXiv preprint*, 2020.

[33] Lars Fischer, Jan-Menno Memmen, Eric MSP Veith, and Martin Tröschel. Adversarial resilience learning – towards systemic vulnerability analysis for large and complex systems. *The Ninth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies*, 9(1):24–32, 2019.

[34] Eric MSP Veith, Nils Wenninghoff, and Emilie Frost. The adversarial resilience learning architecture for ai-based modelling, exploration, and operation of complex cyber-physical systems, 2020.

[35] L. Thurner, A. Scheidler, F. Schafer, J. H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun. pandapower - an open source python tool for convenient modeling, analysis and optimization of electric power systems. *IEEE Transactions on Power Systems*, 2018.

[36] Marita Blank, Timo Breithaupt, Jörg Bremer, Arne Dammasch, Steffen Garske, Thole Klingenberg, Stefanie Koch, Ontje Lünsdorf, Astrid Niesse, Stefan Scherlke, Lutz Hofmann, and Michael Sonnenschein. *Smart Nord Final Report*, pages 21–30. Uni Hannover, 4 2015.

[37] Jeff Heaton. *Introduction to neural networks with Java*. Heaton Research, Inc., 2008.

[38] Christian Spalthoff, Dzanana Sarajlic, Chris Kittl, Simon Drauz, Tanja Kneiske, Christian Rehtanz, and Martin Braun. 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*, pages 1–6. VDE, 2019.

[39] Christian Spalthoff, Dzanana Sarajlic, Chris Kittl, Simon Drauz, Tanja Kneiske, Christian Rehtanz, Martin Braun, Nils Bornhorst, Lars-Peter Lauwen, Jan-Hendrik Menke, Dennis Cronbach, Annika Klettke, Julian Sprey, Tobias van Leeuwen, and Albert Moser. Simbenc – dokumentation, Aug 2019.

[40] Philipp Probst, Marvin N Wright, and Anne-Laure Boulesteix. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3):e1301, 2019.

[41] Eleftherios Spyromitros-Xioufis, Grigorios Tsoumakas, William Groves, and Ioannis Vlahavas. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 104(1):55–98, 2016.

[42] Sepp Hochreiter. Untersuchungen zu dynamischen neuronalen netzen. *Diploma, Technische Universität München*, 91(1), 1991.

[43] Kandethody M Ramachandran and Chris P Tsokos. *Mathematical statistics with applications in R*. Elsevier, 2014.

[44] John Hansen. Using spss for windows and macintosh: analyzing and understanding data, 2005.