

# State Estimation in the Power Distribution System

Dominik Danner  
University of Passau  
dominik.danner@uni-passau.de

Hermann de Meer  
University of Passau  
hermann.demeer@uni-passau.de

## ABSTRACT

In the domain of power distribution network, software that can estimate the grid state using several measurement values as input has been rarely used in the low voltage grid. Such software tools are based on adaptive state estimation methods and their accuracy highly depends on the available input data. Especially, in the low voltage grid which is mostly not monitored at all, the increasing number of controllable high-power loads, such as electric vehicle charging stations or decentralized photovoltaics and battery storage systems, directs the focus to the actual grid state, in particular with regard to the power quality.

This paper discusses how to use machine-learning supported, data-driven state estimation in order to determine the grid state in the low voltage grid irrespective of the grid topology. We compare different input data sampling strategies, like Monte-Carlo, sensitivity analysis, as well as realistic power load profiles with respect to their applicability for training the state estimation. Using correlation analysis, a dependency graph can be modeled, which reflects the important input measurements for each parameter. Furthermore, this paper discusses the accuracy and applicability of different machine learning techniques, such as linear regression, nearest-neighbor and neural networks.

## Keywords

smart grid, state estimation, machine learning, graph theory

## 1. INTRODUCTION

In future, the smart grid will contain thousands of grid state measurement devices in the distribution system and smart metering devices at the households that can provide various different measurement values. Among them, the real and reactive power can be used for billing purposes and the voltage magnitude and angle determine the grid state. Transmitting all these information from all available smart metering device will require a high bandwidth in the future smart grid. Additionally, in case of communication failures the total grid state cannot be monitored anymore. Using state estimation it is possible to determine the grid state based on a subset of the measurements, e.g. with algorithms from literature [1, 2, 3] or the state estima-

tion feature in DIGSILENTs' PowerFactory<sup>1</sup>. Most of these tools perform simulation-based state estimation, which requires information about the grid topology and depends on a Newton-Raphson solver. Especially for smaller *Distribution System Operators* (DSO)s, e.g. municipal utilities, the network maps of their low voltage grids are often not stored in a digital way and sometimes the exact line distances are not even known. In order to overcome this issue but still obtain an accurate grid state, we discuss the usage of machine learning techniques to determine the grid state based on a certain subset of available input measurements in this paper.

## 2. METHODOLOGY

In order to estimate the grid state by machine learning techniques, several important considerations must be taken into account. First, it should be possible to obtain the training data with a reasonable effort. This can be difficult in case of a brown field approach with potentially limited measurement data that may only partially cover subsections of the grid for fractions of time. Second, the input features for the machine learning model need to be chosen carefully and, finally, an appropriate machine learning model need to be trained and validated.

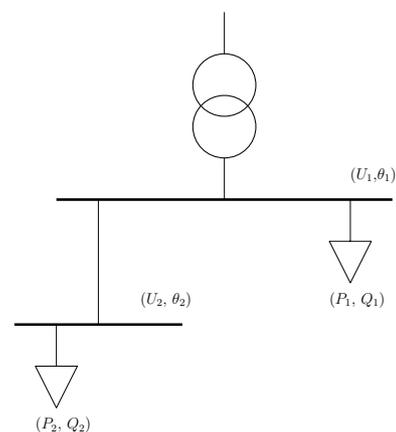


Figure 1: Grid Model

As shown in Figure 1, the low voltage grid consists of lines, buses and loads. Loads can only be attached to buses, which are interconnected via lines. For simplicity, generation com-

<sup>1</sup><https://www.digsilent.de/en/state-estimation.html>

ponents like *photovoltaic* (PV) systems are modeled as negative loads and all loads are assumed to be balanced between the phases. Each load is defined by its real power  $P$  and reactive power  $Q$ , while each bus reveals the voltage magnitude  $U$  and voltage angle  $\theta$  as calculated by a power flow solver or determined by measurement devices.

In the following, training data are generated by sampling  $(P, Q)$  pairs for each load and by running a power solver, e.g. PowerFactory, to obtain the voltage magnitude and voltage angle values. The obtained data set is analyzed variable-by-variable and the other most correlating variables are identified to construct a dependency graph. Based on the graph, machine learning models can be trained for each variable to estimate its value.

## 2.1 Training Data

The quality of the training data plays a very important role on the performance of machine learning models. For that we propose to use three different sampling strategies: (i) Uniform sampling using Monte-Carlo simulation, (ii) one-at-a-time sensitivity analysis and (iii) realistic load profiles.

With Monte-Carlo simulation, the respective input parameters  $P$  and  $Q$  are uniformly sampled, independently from each other between a realistic minimum and maximum value, e.g. taken from realistic load profiles. Furthermore, all  $(P, Q)$  pairs with a power factor less than 0.6 are dropped, for a lower power factor being unrealistic for households. From these finite number of samples, a uniformly sampled  $(P, Q)$  pair is chosen for each load at the low voltage grid.

The one-at-a-time sensitivity analysis fixes uniformly distributed input parameters from a realistic range to all loads except one, where different parameter values from a fixed range are tested. This is repeated until enough data are available for each single parameter to train a machine learning model. Since only one input parameter is changed at the same time, this method likely will reveal the most influencing factors for each parameter.

Considering the realistic context of a low voltage grid, artificially generated input parameters, such as Monte-Carlo sampling or sensitivity analysis are not possible. In order to model real world input data, as it would come from measurements from metering devices, we propose to assign realistic load profiles to each load in the power flow simulation, such as provided by Tjaden et al. in minute or second based resolution [4].

## 2.2 Graph Modeling

After generating the input data for a specific low voltage grid, the data need to be analyzed with regard to their parameter dependencies. We propose to construct a graph that represents the dependencies between the parameters of the low voltage grid. Each parameter ( $P$ ,  $Q$ ,  $U$  or  $\theta$ ) is represented by a node. An edge between two nodes represents a high dependency between these two parameters. An example of such a dependency graph is depicted in Figure 2.

The construction of this dependency graph can be done in two different ways: (i) knowledge-based or (ii) data-driven. Using the grid topology and physical laws, one could extract highly depending parameters and connect them with an edge. Since a machine learning model is to be trained on the gathered data, it need to be ensured that each node has an appropriate number of connections to other nodes.

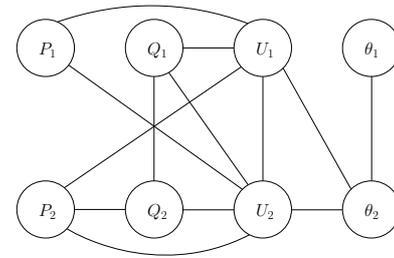


Figure 2: Dependency Graph

Another way to construct the dependency graph is to model it on a data-driven way, e.g. using Pearson correlation coefficient. Each parameter is connected to  $n$  other parameters that yield the highest Pearson correlation. Especially with realistic sampled data it can be expected that at some situations the Pearson correlation is quite high, while no real dependency exists between the nodes. Using one-at-a-time sensitivity analysis seems like a promising sampling strategy for auto generating the dependency graph.

## 2.3 Machine Learning

In the area of machine learning there are many different models that can be applied to the state estimation problem. As known from physics, the power flow calculation is based on a nonlinear system. Hence, using linear regression models will probably not achieve a very high accuracy, unless the grid is mainly operated with a nearly uniform power factor, where the grid can be approximated with linear equations.

In case the training data set is big enough and includes most of the possible grid states, especially high peaks and different combinations of high and low loads at different nodes, nearest-neighbor algorithms, like KNN or *Support Vector Machine* (SVM), could work quite well. In that case, the state estimation will create the estimated state based on the  $k$  most similar trained states.

As a quite different machine learning technique neural networks could be used. Neural networks are inspired by the human brain and embody rules derived from training data. Since the power flow calculation is based on a mathematical model, this class of machine learning algorithms seem to be a promising direction.

## 3. DISCUSSIONS

From the methodology described in Section 2 we expect to obtain local parameter estimations, from which the overall grid state can be inferred. Nevertheless, there are several drawbacks of the proposed approach that will be discussed in the following.

The performance of the proposed data-driven state estimation is expected to highly rely on the quality of the provided training data. In case of one-at-a-time sensitivity analysis, the most influencing parameters can be identified easily and we expect that the dependency graph modeling and machine learning will perform quite well. The accuracy using Monte-Carlo simulation for sampling and nearest neighbor algorithms for learning will probably depend on the data range, whether all important situations are covered. Unfortunately, in a real low voltage grid it is nearly impossible to obtain such ideal data sets. The only realistic sampling strategy (using load profiles) will likely reveal

high correlating connections between parameters, e.g. between the real power of different households due to similar load profiles, which are actually independent variables. As a result, the dependency graph (if not modeled with expert knowledge) will connect wrong parameters and the machine learning model is trained with non reasonable assumptions. Furthermore, we expect that dependency graphs that are built with expert knowledge and consider physical laws will outperform the data-driven construction.

Due to its very local view on the grid - input parameter limited to a subset of all available parameters - the output of each trained parameter model will likely yield inaccuracies. In contrast, this allows a very flexible usage of the state estimation model, since a wide set of different missing values can be tolerated. Estimated output values from some of the parameter models can serve as input to further parameter estimations following the constructed dependency graph. Using Pearson's correlation during construction of the dependency graph is on one hand a reasonable decision, but on the other hand the resulting dependency graph is not directed, because of the symmetry of the correlation coefficient. As a result, at least one of the two parameters of each edge is required to use the trained model for the full grid state estimation. Furthermore, short cyclic paths in the dependency graph limits the overall grid state estimation with missing values.

Because each parameter is estimated on his own, there can be miss-matches between the different parameter models, such that the estimated overall grid state is not reasonable at all or the estimated parameters used as input to other parameter estimation models yield unrealistic output. In order to overcome this issue, a feedback loop could combine the knowledge of all parameter estimation models that incorporate the estimated parameter and optimizes this parameter such that it fits best to all of the models.

Although our methodology requires a high number of small parameter estimation models, the lower number of input parameters leads to a fast model execution, especially when executed on hardware with limited computational power.

## 4. CONCLUSIONS

In this work we discussed the applicability of machine learning models to estimate the grid state in the power distribution system. Furthermore, we proposed an approach to split the state estimation problem to small pieces of parameter estimation models and combine the output of those models to obtain the overall grid state estimation. This paper identified some drawbacks and opportunities of our approach and discussed them in detail.

In future, pseudo measurements, e.g. based on historical data, and short term prediction of load profiles could be integrated to enhance the capability of our model.

## 5. REFERENCES

- [1] Ratmir Gelagaev, Pieter Vermeyen, and Johan Driesen. State estimation in distribution grids. In *2008 13th International Conference on Harmonics and Quality of Power*, pages 1–6, Sep 2008.
- [2] Haibin Wang and Noel N. Schulz. A revised branch current-based distribution system state estimation algorithm and meter placement impact. *IEEE Transactions on Power Systems*, 19(1):207–213, Feb 2004.

- [3] Jianzhong Wu, Yan He, and Nick Jenkins. A robust state estimator for medium voltage distribution networks. *IEEE Transactions on Power Systems*, 28(2):1008–1016, May 2013.
- [4] Tjarko Tjaden, Bergner Joseph, and Volker Quaschnig. Repräsentative elektrische Lastprofile für Wohngebäude in Deutschland auf 1-sekündiger Datenbasis. *HTW Berlin*, pages 1–8, Nov 2015.

## Acknowledgment



This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 713864 (ELECTRIFIC).