

Machine-learning-based Bayesian state estimation in electrical energy systems

Steven de Jongh , Felicitas Müller, Hongfei Li, Boyana Georgieva, Michael Suriyah, Thomas Leibfried

Karlsruhe Institute of Technology (KIT) – Institute of Electrical Energy Systems and High Voltage Engineering (IEH), Karlsruhe, Germany

✉ E-mail: steven.dejongh@kit.edu

Abstract: In many algorithmic applications in electrical power grids, state estimation (SE) represents the first step of a process chain. In SE, sensor measurements are processed to infer the most probable grid state. Classical methods such as weighted least squares (WLSs) based approaches use statistical methods that can be based on sensor noise and erroneous measurements. With these methods, only point estimates are made, which results in a lack of knowledge about prediction uncertainties. In this study, machine-learning-based methods for determining the actual state of the grid are proposed. Bayesian optimisation is applied to find the optimal hyperparameter configurations for neural networks (NNs) for SE tasks. The application of Bayesian inference using Bayesian NNs is proposed, which allows the prediction of point estimates as well as uncertainty intervals for the system states. The advantages of using Bayesian approaches in comparison to classical SE methods like WLS are shown.

1 Introduction

The increasing number of renewable energy generation plants, such as wind turbines and photovoltaic generators and the addition of new consumers, for example, electric vehicles and heat pumps, represent a major challenge for the operation of future electrical grids. The weather-dependent and therefore volatile nature of renewable generation plants makes accurate infeed forecasts for these plants challenging. In addition, many of the renewable plants feed decentralised into the local distribution grids, leading to new load flow situations. Furthermore, the number of loads in the distribution grid with high-power consumption will increase due to the introduction of charging equipment for battery electric vehicles and the coupling of the heat, gas and electricity sectors. The combination of these factors leads to higher stresses on the grid equipment. In order to be able to guarantee conformity with grid codes and prevent damage to grid equipment, the monitoring of the actual grid state is of great importance. Based on the current system state, measures can be taken to comply with grid codes, optimise equipment scheduling and be used for future expansion decisions.

State estimation (SE) is the process of inferring the values of the system states (i.e. the voltages at all nodes of the grid) using a limited number of measured data at certain sensor positions in the system [1]. Although SE is a widely adopted concept in electrical transmission systems, the applications of SE in distribution systems are still new. Different publications focused on the application of weighted least squares (WLSs) based SE with noisy measurements and limited observability [2]. Applications range from single-phase to three-phase electrical systems, as well as coupled systems [3]. Different approaches using various algorithms have been proposed [4]. All of the statistical methods are based on assumptions regarding the distributions of measurement errors as well as physical models of the respective grid. Data-based methods on the other hand make direct use of the sensor data obtained offering several advantages since they do not directly rely on an algebraic description of the physical grid. Furthermore, the observability of the entire system does not have to be ensured manually by adding pseudo-measurements. Furthermore, cross-correlations between measurements at different nodes can be learned from the data. In addition, joint estimation of

profiles as well as the use of exogenous parameters (weather, time, behavioural profiles) can be used.

SE using neural networks (NNs) has been performed in various papers [5–7] used a hybrid model based on machine learning and statistical techniques to improve the performance of SE. In Menke *et al.* [8], feed-forward NNs are used for distribution system SE. It is shown, that the proposed ANN-based approach is able to outperform classical WLS methods. Recent approaches include the usage of hybrid models [9] and applications of pruned NNs [10], increasing efficiency and accuracy on SE tasks.

In this paper, two aspects extend the consideration of SE using NNs. First, Bayesian optimisation (BO) is used to determine optimal hyperparameter configurations. In addition, Bayesian NNs (BNNs) are applied as a method that is not based on point estimates but includes uncertainties in the estimates and is, therefore, better suited for risk assessments during grid operation.

2 Methodology

2.1 State estimation using NNs

SE can be applied to estimate the current system state of an electrical grid. Either all complex node voltages or all complex branch currents can express the system state. The state of the system \mathbf{x} is defined as $\mathbf{x} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]^T$ and is a $N \times 1$ vector, where N is the number of nodes in the system. The system is fully defined when all complex node voltages \mathbf{v}_i are known, which means that all power flows can be calculated. When measuring in real grids, the measurements can be subject to measurement errors. Equation (1) shows the generic measurement equation.

$$\mathbf{z}_t = \mathbf{h}(\mathbf{x}_t) + \boldsymbol{\varepsilon}_t \quad (1)$$

\mathbf{z}_t is the measurement vector at timestep t . \mathbf{h} is the function that maps the states to measurements. For measurements of power, this mapping is non-linear and non-convex making the resulting system unsolvable using a matrix inversion. $\boldsymbol{\varepsilon}_t$ is the measurement error vector. SE has the goal to find the system state that is most likely, given the measurements. The SE problem can be simplified

to a regression problem, where the goal is to find a function $f(z_i)$ that maps the noisy measurements to an approximation of the actual system state \hat{x}_i . When applying a machine-learning algorithm to this problem, the function that maps the measurements to approximated states has to be found by training using historic data. Equation (2a) shows the supervised learning problem that is solved, minimising the mean-squared error (MSE) shown in (2b).

$$\mathbf{X}_{\text{train}} \leftarrow f_{\theta}(\mathbf{Z}_{\text{train}}) \quad (2a)$$

$$\text{MSE} = \frac{1}{M \times T} (\mathbf{X}_{\text{train}} - \hat{\mathbf{X}})^2 \quad (2b)$$

$\mathbf{X}_{\text{train}}$ is the matrix containing T samples of historic true system states and has a dimension of $N \times T$. Correspondingly $\mathbf{Z}_{\text{train}}$ is the $M \times T$ matrix, where M is the number of measurements per sample. θ is the parameter of the function approximator that has to be learned. In this paper, NNs are used as a function approximator. The goal is to train a generic parameterised function f_{θ} , which maps from measurements to approximated states. Equation (3) shows a function represented by a two-layer NN.

$$y_{l+2} = \sigma_{l+1}(\mathbf{W}_{l+1} \sigma_l(\mathbf{W}_l y_l + \mathbf{b}_l) + \mathbf{b}_{l+1}) \quad (3)$$

The output of layer l (y_{l+1}) is calculated by multiplying a weight matrix (\mathbf{W}_l) by the input of the layer (y_l) and adding the bias \mathbf{b}_l . Afterwards, the resulting vector is passed through a differentiable activation function σ_l . By stacking multiple layers, more complex structures emerge.

2.2 Extension to Bayesian case

Classical NNs can be expressed as a probabilistic model $P(y|x, \mathbf{W})$, where based on the input data x , the output y is predicted in a regression task. In this case, the weights can be found by maximum likelihood estimation [11]. Equation (4) shows the objective function that is solved by gradient descent.

$$\mathbf{W}_* = - \underset{\mathbf{W}}{\text{argmin}} \sum_i \log P(y_i|x_i, \mathbf{W}) \quad (4)$$

For Gaussian distributions P , (4) is the minimisation of the MSE. This shows that training a NN using MSE is equivalent to applying WLS-SE. Since both methods are frequently used statistical models, only point estimates are inferred. When performing Bayesian inference with NNs, each weight of the NN is a distribution with a mean value μ and standard deviation σ . The goal is to find the posterior distribution $P(\mathbf{W}|\mathcal{D})$ given the training data \mathcal{D} leading to the objective in (5).

$$J(\mathcal{D}, \theta) = \text{KL}[q(\mathbf{W}|\theta) \| P(\mathbf{W})] - \mathbb{E}_{q(\mathbf{W}|\theta)}[\log P(\mathcal{D}|\mathbf{W})] \quad (5)$$

The first part of (5) minimises the Kullback-Leibler (KL) divergence between approximated and true posterior. Using variational inference and Bayes by backprop [11], it is possible to find the BNN parameters that minimise this objective function. When applying the NN in a regression task, the network not only predicts the given point estimates but also the expected uncertainties of its prediction.

2.3 Benchmark grid and measurement setup

To test the machine-learning-based SE algorithms, the distribution grid, shown in Fig. 1, is used [12]. The 400 V radial feeder consists of 15 nodes with point-of-common-coupling at node 0 at 20 kV. At each node, a load is connected simulating the industrial and household loads. At nodes 1, 2, 3, 5, 7, 8, 11 and 13 photovoltaic generators are installed. As measurement equipment, phasor measurement units (PMUs) are assumed which measure the voltage magnitude, voltage phasor, active- and reactive-node

power at given time intervals. Standard load and generation profiles are used for data generation.

2.4 Data generation and network training

The training and testing data are generated using power flow calculations for each 15-min time step of a year resulting in 35,136 data points. To simulate the measurement noise of the PMU devices, Gaussian noise with $\sigma_{P,Q} = 1.0\%$ and $\sigma_{v,\theta} = 0.366\%$ is applied. An 80/20 train-test split is applied. Table 1 shows the different PMU configurations being considered.

The supervised training of the NNs is done using the training data. The input of the NN consists of the PMU measurements depending on the scenario (Table 1). The target consists of the voltage magnitudes and voltage angles of all nodes in the system. Fig. 2 shows an example setup for a three bus system with two PMUs

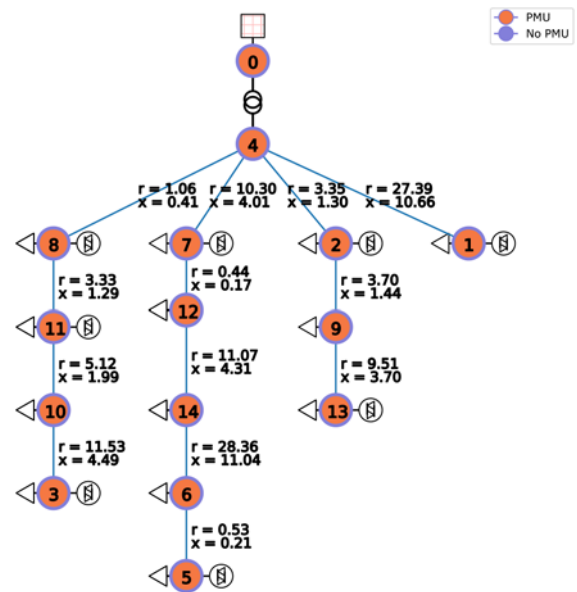


Fig. 1 Benchmark distribution grid under consideration; resistance (r) and reactance (x) in $\text{m}\Omega$

Table 1 Considered sensor placements

Scenario	PMU setup	Penetration τ
0	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14}	1.0
1	{1, 2, 3, 5, 7, 8, 13}	0.47
2	{1, 3, 7, 10, 12}	0.33
3	{1, 2, 7, 8}	0.27
4	{1, 3, 5, 13}	0.27
5	{4, 5}	0.13

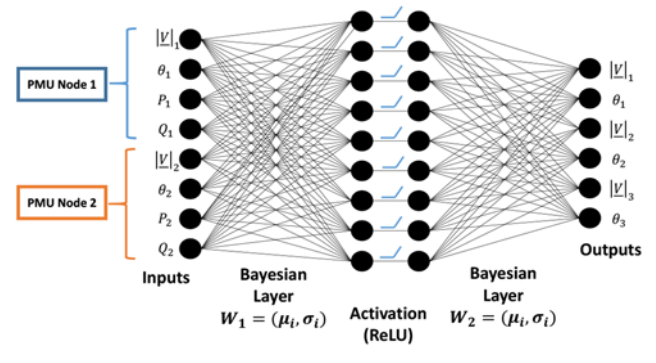


Fig. 2 Example 2 layer BNN for three-bus system

Table 2 Optimised hyperparameters

Hyperparameter	Optimised value
training epochs	9555
num layers	4
neurons layer 1	94
neurons layer 2	78
neurons layer 3	62
learning rate	2.1×10^{-4}
activation 1	LogSigmoid
activation 2	exponential linear
activation 3	exponential linear
bias	off

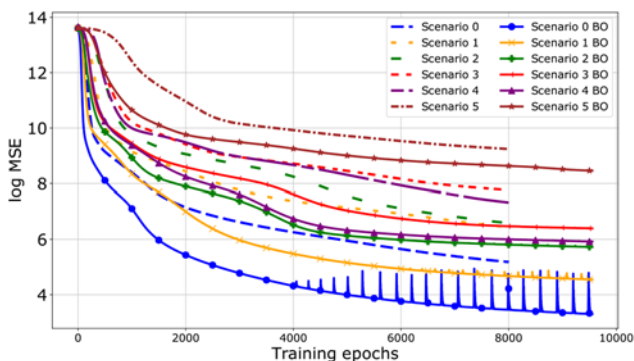
installed at nodes 1 and 2. Node 3 is not equipped with a PMU, hence only appears as a target. Although Fig. 2 considers two Bayesian layers, an arbitrary number of layers, neurons per layer and activation function are possible. The optimisation of the hyperparameters is achieved using 100 runs of BO where the objective is to find the best hyperparameter setting given the evaluated model performance of previously trained networks.

3 Results

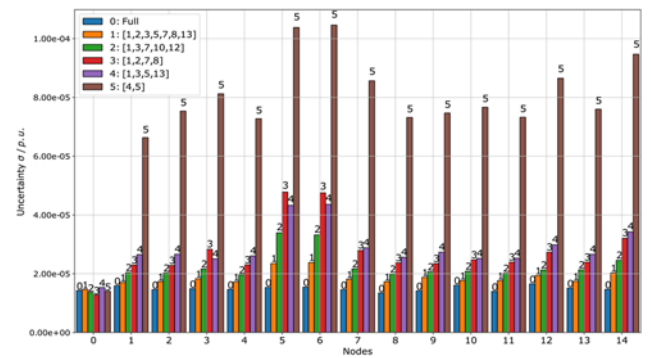
First, classical NNs are used for SE. These carry out point estimates of the system states. The initial setup consists of NNs that are trained for 8000 epochs using the Adam optimiser with a learning rate of 0.0001. Four hidden layers with 1.2 times the inputs as neurons are used. Each layer, besides the output layer, uses the rectified linear unit as activation. BO is applied for 100 iterations to optimise the hyperparameters. Table 2 shows the optimised values for the grid with full sensor equipment. After optimisation, an NN with three hidden layers is obtained. The number of neurons decreases for each layer. The optimal bias variance trade-off is achieved when training for 9555 epochs.

Fig. 3 shows the training procedure for the six different scenarios of Table 1. It can be seen, that after hyperparameter optimisation the training procedure is faster and reaches a lower training MSE.

In Table 3, the performances on the testing data of the NN with the initial and optimised setups are shown. It can be seen, that through

**Fig. 3** Training of NN for different sensor scenarios before and after BO**Table 3** Sensor scenarios testing

ID	PMU setup	Testing MSE (initial setup)	Testing MSE (BO)
0	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14}	1.439×10^{-5}	0.624×10^{-5}
1	{1, 2, 3, 5, 7, 8, 13}	3.688×10^{-5}	1.265×10^{-5}
2	{1, 3, 7, 10, 12}	5.449×10^{-5}	2.494×10^{-5}
3	{1, 2, 7, 8}	9.164×10^{-5}	3.155×10^{-5}
4	{1, 3, 5, 13}	8.201×10^{-5}	2.627×10^{-5}
5	{4, 5}	40.628×10^{-5}	7.824×10^{-5}

**Fig. 4** Predicted voltage uncertainty of voltage magnitude ($t=842$) for different sensor scenarios

hyperparameter optimisation an increase in performance is achieved for all sensor placement scenarios. Furthermore, the error increases with lower sensor penetration. When the number of PMUs is lower than the number of grid branches, the estimation error increases sharply, motivating optimised sensor placements in the planning phase.

By using BNN for SE, not only the point estimates of the states but also uncertainty estimates can be made. Fig. 4 shows the uncertainty estimates for an example time step. It can be seen, that the uncertainty increases when the number of installed PMUs decreases. The highest uncertainty is predicted for nodes 5 and 6, as they are connected to the rest of the grid via a high-impedance power line. In scenario 4 (PMUs at end of feeder), the uncertainties for the nodes that are at the end of the respective feeders (3, 5 and 6) is better compared to scenario 3 (PMUs at beginning of feeder) where the results are better for the rest of the nodes. The predictions for node 0 are similar for all scenarios since this is the slack node of the system being assumed to have a fixed voltage magnitude and angle.

4 Conclusion

It could be shown in this paper that the combination of statistical and machine-learning methods can lead to very good results in SE of electrical distribution grids with low sensor equipment. First, BO was used to optimise the hyperparameters of classical NN. Then, the NN is extended to BNNs, which allows us to infer the uncertainty of the prediction. The uncertainties can be used in subsequently used algorithms that make use of this information and thus guarantee a more reliable operation compared to point predictions. A conceivable application could be the use of stochastic programming, where the prediction of the uncertainty can be used directly to provide certain guarantees of the optimisation results.

In the future, BNN can be combined with geometric deep learning techniques to allow faster training times and inductive behaviour by making use of the grid topology. Furthermore, the usage of exogenous inputs and dynamic SE techniques, i.e. through recurrent NNs, is a promising field of research.

5 References

- Monticelli, A.: 'State estimation in electrical power systems: a generalized approach' (Springer Science & Business Media, USA, 1999, 1st edn.)
- Liu, J., Tang, J., Ponci, F., et al.: 'Trade-offs in PMU deployment for state estimation in active distribution grids', *IEEE Trans. Smart Grids*, 2012, 3, (2), pp. 915–924
- Pfeifer, M., Zimmerlin, M., Schwarzendorfer, M., et al.: 'Weighted least squares state estimation for coupled power and gas distribution networks'. 53rd IEEE Universities Power Engineering Conf., Glasgow, UK, 2018, pp. 1–6
- Dehghanpour, K., Wange, Z., Wang, J., et al.: 'A survey on state estimation techniques and challenges in smart distribution systems', *IEEE Trans. Smart Grid*, 2019, 10, (2), pp. 2312–2322

- 5 Pertl, M., Heussen, K., Gehrke, O., *et al.*: 'Voltage estimation in active distribution grids using neural networks'. 2016 IEEE Power and Energy Society General Meeting, Boston, 2016, pp. 1–5
- 6 Abdel-Nasser, M., Mahmoud, K., Kashef, H.: 'A novel smart grid state estimation method based on neural networks', *Inter. J. Inter. Multimedia Artif. Intell.*, 2018, **5**, pp. 92–100
- 7 Wu, J., He, Y., Jenkins, N.: 'A robust state estimator for medium voltage distribution networks', *IEEE Trans. Power Syst.*, 2013, **28**, (2), pp. 1008–1016
- 8 Menke, J., Bornhorst, N., Braun, M.: 'Distribution system monitoring for smart power grids with distributed generation using artificial neural networks', *Int. J. Electr. Power Energy Syst.*, 2018, **113**, pp. 472–480
- 9 Zhang, L., Wang, G., Giannakis, G.: 'Real-time power system state estimation and forecasting via deep neural networks', *IEEE Trans. Signal Process.*, 2019, **67**, (15), pp. 4069–4077
- 10 Zamzam, A., Sidiropoulos, N.: 'Physics-aware neural networks for distribution system state estimation', *IEEE Transaction on Power Systems*, 2020, **35**, (6), pp. 4347–4356
- 11 Blundell, C., Cornebise, J., Kavukcuoglu, K., *et al.*: 'Weight uncertainty in neural networks', *Proceedings of the 32nd International Conference on Machine Learning*, 2015, vol. 37, pp. 1613–1622
- 12 Meinecke, S., Bornhorst, N., Braun, M.: 'Power system benchmark generation methodology'. *Conf. on Sustainable Energy Supply and Energy Storage Systems*, Hamburg, Germany, 2018