Application of Synthetically Trained Neural State Estimation to Real-World Distribution Grids

A dissertation approved for the academic degree of

Doctor of Engineering (Dr.-Ing.)

at the

Faculty of Electrical Engineering
and Information Technology

of

TU Dortmund University

by

Thomas Oberließen, M.Sc.
from Paderborn

Dortmund 2025

Supervisor:	Univ.-Prof. Dr.-Ing. habil. Christian Rehtanz
Co-Advisor:	Univ.-Prof. Dr. rer. nat. Sebastian Lehnhoff
Day of the oral examination:	16 October 2025

Abstract

The integration of renewable energies and sector coupling advances lead to an increasingly complex supply environment in distribution grids. Whereas active grid monitoring and operational control were historically uncommon in distribution grids, their significance is now markedly increasing. Recent research points, among others, to supervised learning-based approaches as an effective solution for distribution system state estimation. A fundamental limitation of those approaches is the reliance on synthetic training data for training and testing. This thesis models a synthetic variant of a real distribution grid and investigates the transferability of a synthetically trained model to the measurement data of the real grid. Real-world measurement data inevitably contains errors and missing signals, presenting a significant challenge for state estimation. Thus, the thesis investigates the detection and reconstruction of anomalies within the measurement data. Furthermore, the limited transparency of machine learning models hinders their adoption in the field. Consequently, an extension of the supervised learning model to a stochastic state estimation that quantifies estimation uncertainty and delivers estimation ranges is proposed.

Kurzfassung

Die Integration erneuerbarer Energien und Fortschritte der Sektorkopplung führen zu einer zunehmend komplexen Versorgungssituation in Verteilnetzen. Während aktive Netzüberwachung und Betriebsführung in Verteilnetzen historisch unüblich waren, nimmt ihre Bedeutung nun deutlich zu. Aktuelle Forschung verweist, unter anderem, auf Ansätze basierend auf überwachtem Lernen als eine effektive Lösung für die Verteilnetz-Zustandsschätzung. Eine fundamentale Einschränkung dieser Ansätze ist die Abhängigkeit von synthetischen Trainingsdaten für das Training und Testen der Modelle. Diese Arbeit modelliert eine synthetische Variante eines realen Verteilnetzes und untersucht die Übertragbarkeit eines synthetisch trainierten Modells auf die Messdaten des realen Netzes. Reale Messdaten enthalten unweigerlich Fehler und fehlende Signale, was eine signifikante Herausforderung für die Zustandsschätzung darstellt. Daher untersucht die Arbeit die Detektion und Rekonstruktion von Anomalien in realen Messdaten. Darüber hinaus behindert die begrenzte Transparenz von Machine-Learning-Modellen ihre Anwendung in der Praxis. Folglich wird eine Erweiterung des Modells des überwachten Lernens zu einer stochastischen Zustandsschätzung vorgeschlagen, die die Schätzunsicherheit quantifiziert und Schätzbereiche liefert.

Acknowledgment

"I don't know where I am going, but I know exactly how to get there."

– Boyd Varty

Researching and writing a phd thesis is not only an intellectual endeavour but also a personal development journey. During my time at the Institute of Energy Systems, Energy Efficiency and Energy Economics I had the opportunity to deepen my technical expertise, collaborate with inspiring colleagues, and grow both as a researcher and as a person. First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Dr.-Ing. Christian Rehtanz, for his invaluable guidance, encouragement, and expertise throughout my doctoral studies. His scientific insight, constructive feedback, and constant support have been instrumental in shaping the direction and quality of this work. I am especially thankful for the freedom and trust he afforded me to explore my research interests while always providing thoughtful advice when needed.

I would also like to thank all my colleagues at the Institute of Energy Systems, Energy Efficiency and Energy Economics for creating an inspiring and collaborative research environment. The stimulating discussions, shared challenges, and many moments of laughter made my time at the institute both professionally rewarding and personally enjoyable. Special thanks go to my research group for their support, cooperation, and friendship over the years.

My appreciation extends to our project partners and collaborators for their valuable input, constructive exchanges, and the opportunity to translate academic research into practical applications. Their perspectives greatly enriched this thesis.

On a personal note, I am profoundly grateful to my family and friends for their unconditional love, patience, and encouragement throughout this journey. Their belief in me kept me motivated through both the exciting and the difficult phases of this work.

Finally, I would like to thank everyone who, in one way or another, contributed to the completion of this dissertation. This achievement would not have been possible without your support.

Contents

1 Introduction

The ongoing energy transition is fundamentally transforming Distribution Grids (DGs). In 2024, 58 % of all electricity in Germany was generated from renewable energies [1]. Furthermore, advances in sector coupling lead to an increasing amount of loads like electric vehicles and heat pumps [1] and energy management systems are being deployed to control prosumer devices effectively [2]. A significant part of these changes occur in DGs and can lead to unprecedented stress on the grid due to high generation and load peaks and an overall increase in the complexity of the grid [3].

These developments necessitate enhanced control mechanisms for distribution grids, a need that has been acknowledged by regulatory measures. For instance, the German Energy Industry Act (EnWG) § 14a explicitly obligates grid operators to control specific loads in case of an observed or forecasted grid congestion [4]. However, a fundamental challenge emerges: grid operators cannot address problems they cannot detect. All grid operation and automation measures for DG are downstream of reliable DG monitoring systems, which rely on a sufficient amount of measurement infrastructure [5]. Despite long-standing awareness of the need for enhanced measurement infrastructure, initiatives to increase measurement penetration in distribution grids, particularly the smart meter rollout, have not progressed as rapidly as needed [6]. Consequently, distribution grids currently operate with limited measurement infrastructure, resulting in largely unknown grid states across most distribution grids.

State Estimation (SE) is the process of estimating the state of an electrical grid based on available measurements. While conventional state estimation methods require a mathematically observable grid, this requirement often cannot be met in DGs due to their limited measurement infrastructure. To address this challenge, Distribution System State Estimation (DSSE) approaches center around estimating the state of unmeasured electrical buses using only a limited set of measurements [7]. While there are different approaches for DSSE, increasing attention is paid to the usage of Machine Learning (ML) based supervised learning techniques [8].

Supervised Neural State Estimation

The fundamental principle of supervised learning-based DSSE is the usage of labeled data samples to train a model for specific tasks [9]. In the context of DSSE, this

means that a model is trained to estimate the states of unmeasured buses based on the available measurements within the power system. During the training process, these estimations are compared against the true states of the unmeasured buses, as illustrated in Figure 1.1. The difference between the estimated and true states serves as the basis for adjusting the model parameters, ultimately improving its estimation performance. When using Artificial Neural Network (ANN) as the underlying model for SE, the process can be refered to as Neural State Estimation (NSE). However, this approach faces a fundamental challenge: by definition, the actual states of unmeasured buses are unknown in real-world applications. Consequently, the very data one tries to estimate with the algorithm, is the data that is needed to train the DSSE model.

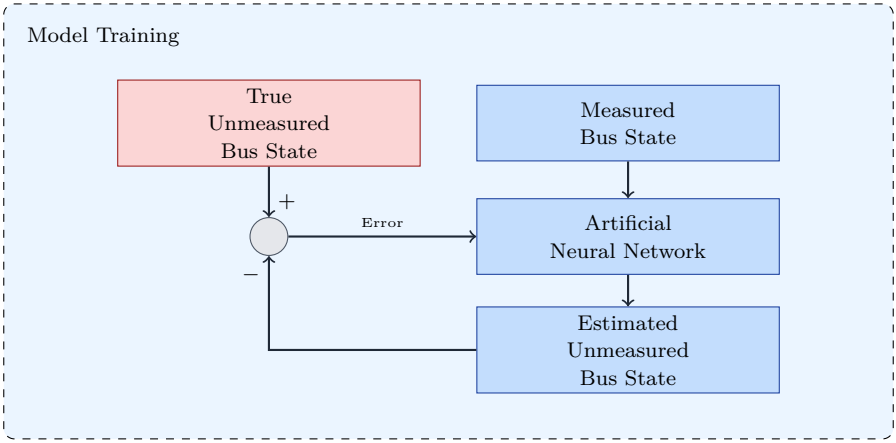


Figure 1.1: Simplified process of supervised learning for DSSE.

To circumvent the problem, current literature approaches rely mostly on synthetic data for model development and validation. Most publications in this field utilize standardized IEEE test grids as their foundation, complemented by generated time series data that simulate bus load and generation patterns [10]–[13]. By conducting power flow calculations using these synthetic load and generation profiles mapped onto the electrical grid topology, researchers can create comprehensive datasets containing complete grid states suitable for model training. The general approach is

outlined in Figure 1.2. The validation of these models typically follows standard machine learning practices, where model performance is evaluated using a separate test dataset that was explicitly excluded from the training process. While this approach enables the development and testing of supervised learning methods for DSSE in this artificial context, it raises important questions about the applicability of the methodolgy for real world distribution grids.

An increase in digitization of distribution grid data can be observed. This trend is reflected in the growing research into digital twins of power systems [14]. A sufficiently precise digital representation of the actual system might be the missing bridge between synthetic model training and real application, which is the general context of this dissertation. If the digital representation of the model can be leveraged to produce a system state representation that is reasonably similar to the real system behavior a synthetically trained system might be transferable to its field application. Information on today's distribution grids is mostly sparse, especially with respect to information about the load and generation systems connected to the grid. Thus, the approach should be evaluated with a realistic data set available to grid operators. The main research question synthesizes these considerations and is:

Research Question 1: *Is a NSE model trained on a synthetic representation of a real grid transferable to field operation given our limited information about real distribution grids?*

When dealing with real measurement data instead of synthetically generated signals, data quality challenges arise. On top of measurement noise, measurements can for example be corrupted or missing. Corrupted data can lead to a degradation of estimation performance. Furthermore, missing signals pose a problem for models, like ANNs, that expect a fixed-size input vector. Thus, detecting anomalies and reconstructing missing signals is of considerable importance for employing NSE in the field and the theme for the second research question, which is:

Research Question 2: *Can synthetically trained ANNs be leveraged for anomaly detection and signal reconstruction of field measurements in the context of NSE?*

The process of estimating system states that are not explicitly measured is inherently imprecise. This is especially true when used in grids with very sparse measurement infrastructure. Additionally, there is a common concern surrounding the limited transparency of ANNs due to the difficulty of interpreting how the models arrive at

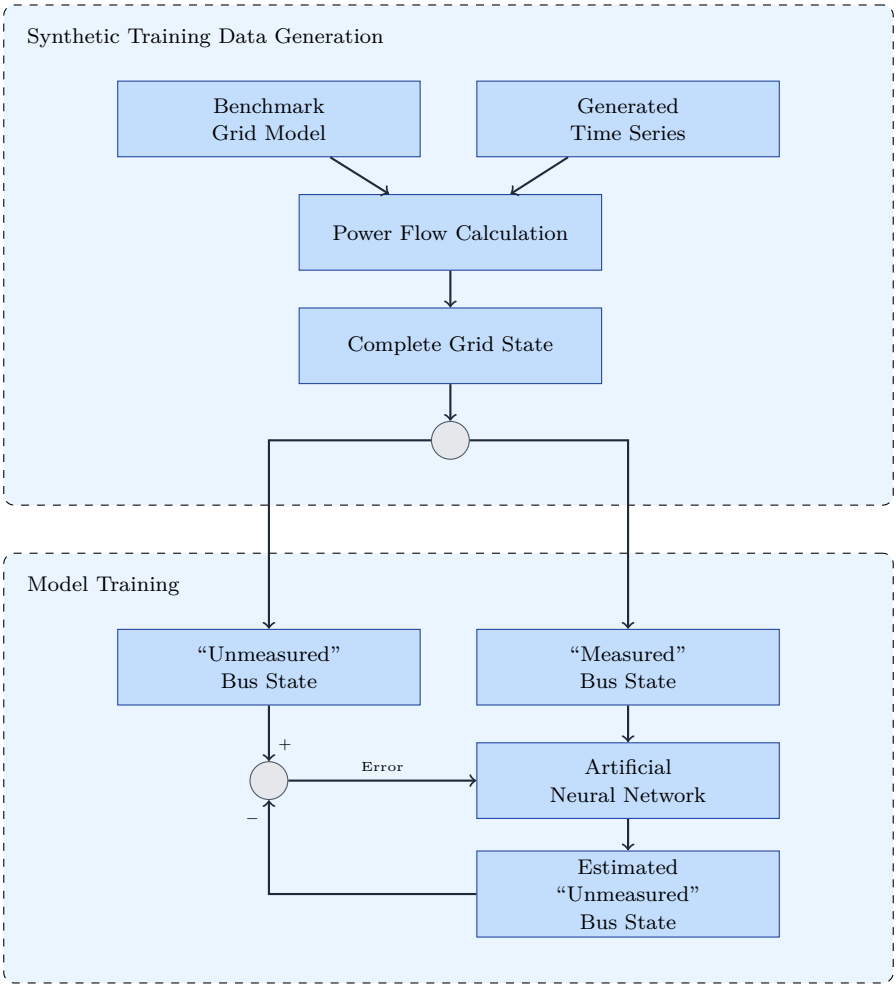


Figure 1.2: Conceptual overview of NSE using synthetic data.

their specific solution. These challenges highlight the need for methods that not only provide state estimates but also quantify the associated uncertainty. Given a method for uncertainty quantification, the process of state estimation can be extended to include prediction ranges in a stochastic DSSE. This leads to the third research question:

Research Question 3: *How can the estimation uncertainty of NSE models be quantified and used for stochastic SE?*

Outline and Contributions

This thesis is structured as follows: Chapter 2 describes the fundamentals of the DSSE. The conventional SE approaches are outlined and the different types of measurements available for a DSSE are described. Furthermore, the basics of superverised ML are described and how these approaches are used in the context of a DSSE. The DSSE has certain associated challenges. Some of them are specific to the problem domain itself and some are specific to the application of supervised methods in the DSSE context. Those are described and form the basis for implementation decisions of the approach of this thesis.

Chapter 3 is centered around a real DG that forms the data basis for the following chapters. The grid model itself and the corresponding measurements from the field are investigated. Based on this model a synthetic grid model including the load and generation systems is created. Using the synthetic grid model a complete synthetic data set is simulated and validated against the measurements from the field. By generating a synthetic complete data set supervised learning models can be trained on a representation of an actual power system.

In summary, the contributions of Chapter 3 are

- an approach for modeling a distribution grid for synthetic data generation;
- a demonstration of its ability to generate similar data distributions to its real-world counterpart;

Real measurement data can be missing or corrupted due to various reasons. Since dealing with these cases is essential for a robust DSSE, Chapter 4 investigates the usage of Autoencoders for anomaly detection. Furthermore, the same model is leveraged for reconstructing anomalous signals. The model for anomaly detection and signal reconstruction is shown to be effective on the synthetic data set. Additionally, the transferability of these models to the actual field measurements is analyzed.

In summary, the contributions of Chapter 4 are

- a demonstration of supervised-learning based anomaly detection transferability to field application;
- a supervised-learning based signal reconstruction approach for anomalous data and its transferability to field application;

Chapter 5 investigates the actual NSE implementation and compares it to the more conventional Weighted Least Squares (WLS) method. After training and evaluating the methods on the synthetic data set, the field measurements are used to analyze performance in a field setting. The robustness of the NSE is investigated and a mechanism for fine tuning is proposed and evaluated. To increase the transparency of the model estimation uncertainty the usage of dropout layers is proposed for quantifying the model uncertainty. The technique is extended to incorporate estimation intervals given the historical model errors to deliver a probabilistic state estimation.

In summary, the contributions of Chapter 5 are

- NSE field application transferability analysis;
- NSE fine-tuning technique for increased robustness against load and generation changes;
- Monte Carlo Dropout for DSSE uncertainty estimation;
- probabilistic DSSE based on Monte Carlo Dropout;

In Chapter 6 a structural overview of the resulting components and their integration into an overall system is given. Furthermore practical implementation considerations are outlined. Chapter 7 concludes the thesis by summarizing the results and answering the research questions that are posed. Finally, future research directions are stated in the corresponding outlook section.

2 Fundamentals

This chapter provides the fundamental concepts and theoretical background of conventional state estimation approaches and modern neural network-based methods. The chapter begins by examining grid monitoring in distribution systems, including the essential aspects of Distribution System State Estimation (DSSE) use cases and formalizes the state estimation problem. It then explores different types of measurements and their accuracy definition. The second to last section introduces the principles of supervised learning and neural networks and their application to the DSSE, before concluding with a comprehensive overview of current challenges and requirements in DSSE implementation.

2.1 Grid Monitoring in Distribution Systems

2.1.1 Congestion Management and DSSE Application

The DSSE is an operational monitoring tool. Its fundamental purpose is the establishment of grid observability, by estimating the system state given a set of field measurements, enhanced by additional data sources. The system state is defined by the complex bus voltages and the resulting current flows along the power lines [7]. As such it forms the basis for any operation and control mechanism in distribution grids [15]. One central operational task that builds upon the grid observability established by the DSSE is congestion management.

During congestion management limit violation of the electrical infrastructure is monitored and alleviation measures are activated to prevent equipment damage [16]. The two fundamental types of congestion include voltage and current congestions. A voltage congestion occurs if a bus voltage falls outside the range of plus or minus 10% of nominal voltage as referenced in the applicable DIN EN 50160 norm [17]. Current congestions occur when power flows exceed the rated current capacity of the power line. By monitoring the system state the DSSE can detect congestions and trigger operational measures. One of the most common measures for voltage control is the usage of on-load tap-changers. These can control voltage levels by adjusting the transformers tap position as shown in [18] and [19]. Grid reconfiguration measures can be employed if switches are present in the Distribution Grid (DG), as investigated in [20]. Furthermore, reactive power can be managed by leveraging Renewable Energy Sources (RES) as shown in [21]. More active managing of RES and controllable loads plays an incresing role in congestion management. The

recent ğ 14a ENWG obligates DG operators to put a DSSE in place in the near future to identify grid congestions. In case of a congestion certain types of load, like heat pumps and non-public electric vehicle charging stations should be actively controlled by the grid operator. More broadly a lot of publication investigate the effectiveness of flexibilities, controllable load and generation, on grid congestion management [22]–[24]. On a longer time horizon the detection of grid congestions as well as an establishment of infrastructure utilization is essential for refined grid extension planning.

2.1.2 State Estimation Problem Formulation

Physical State Description

The state of the power system can be described by the collection of all complex bus voltages $x_i = [v_i, \delta_i]$, made up of the voltage magnitude v_i and the voltage angle δ_i of the indivdual buses i. Values in this section are given in p.u.. We can therefore define the system state as a state vector $x = [x_1, x_2, ..., x_n]$ with $i = 1, 2, \ldots, n$ where n is the number of buses.

The different types of measurement variables can be related to that state vector. Active and reactive power can be described by the power injection equation

$$\begin{cases} P_i = v_i \sum_{j=1}^{n} y_{ij} v_j \cos\left(\delta_i - \delta_j - \theta_{ij}\right), & \forall i, \\ Q_i = v_i \sum_{j=1}^{n} y_{ij} v_j \sin\left(\delta_i - \delta_j - \theta_{ij}\right), & \forall i \end{cases} \tag{2.1}$$

where P_i and Q_i are the active power and reactive power injections at bus i. $v_{i(j)}$ is the bus voltage magnitude, $\delta_{i(j)}$ the voltage angle at bus $i(j)$. $y_{ij}\angle\theta_{ij}$ is the ijth element of the admittance matrix of the system.

Likewise line currents can be expressed as a function of the system's complex voltages, given

$$I_{ij} = \sqrt{(g_{ij}^2 + b_{ij}^2)(v_i^2 + v_j^2 - 2 v_i v_j \cos\theta_{ij})} \tag{2.2}$$

where I_{ij} is the branch current magnitude between buses i and j and $V_{i(j)}$ the voltage magnitude of bus $i(j)$. g_{ij} is the conductance and b_{ij} is the susceptance with respect to the ijth entry of the systemes admittance matrix [7].

Measurement Function

The previously described system state is never known but can only be observed by measurement devices. All measurements of the system can be composed into a measurement vector $z = [z_1, z_2, \ldots, z_m]$, where m is the number of measurements. Given that all measurements contain some level of measurement noise, the measurement vector can be represented by

$$z = h(x) + e \tag{2.3}$$

The non-linear function $h(x)$ expresses the measurements in terms of the grids state variables using the grid equations and e describes the measurement noise. The overall state estimation task is to estimate the state vector x that best satisfies Equation 2.3 [7], [25].

Grid Obeservability

An important characteristic as it relates to grid calculations and their various algorithms is the observability of the grid. The observability characterizes the amount of measurements m in relation to the grids' amount of buses n. Mathematically this can be described using the measurement redundancy η where

$$\eta = \frac{m}{2 \times n - 1} \tag{2.4}$$

Given the redundancy η the observability can be classified as overdetermined, exactly determined and underdetermined as depicted in Figure 2.1.

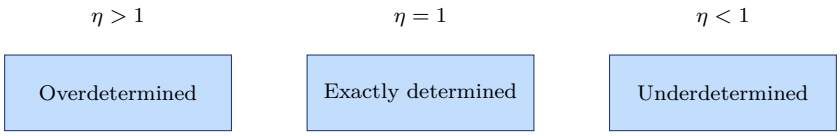


Figure 2.1: Classification of grid observability based on measurement redundancy [22].

Conventional state estimation approaches rely on an η value that is greater than one and oftentimes values greater than two are recommended [7], [25]. Due to the limited measurement infrastructure in DG the challenge of DSSE is handling redundancies smaller than one.

2.1.3 Conventional State Estimation

Weighted Least Squares

In transmission system the State Estimation (SE) is traditionally solved using the Weighted Least Squares (WLS) algorithm. In order to run the algorithm, the system needs to be observable so the measurement redundancy η needs to be greater or equal to one. It uses an iterative approach to minimize the least square of the error e. Given Equation 2.3 this results in the following objective function

$$\min J(x) = (z - h(x))^T W (z - h(x)) \tag{2.5}$$

where $W = R^{-1}$ represents the measurement weight matrix with which the measurements are weighted by the inverse of their variance. For the mathematical derivation of the formula reference [25].

$$W = \begin{bmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_m^2 \end{bmatrix} \tag{2.6}$$

The optimal solution can be found using the optimality condition of $J(x)$:

$$\frac{\partial J(x)}{\partial x} = -2H^T(x)W(z - h(x)) = 0 \tag{2.7}$$

$H(x)$ represents the non-linear Jacobian matrix which constitutes the partial derivatives of $h(x)$ with respect to x. The resulting system of equations is typically solved using iterative procedures. One common approach is to use Newton's method, using the first terms of the Taylor series [25]. Using this method the following gain matrix G is obtained, which represents the sensitivity of the measurements to changes in the state variables

$$G(x_\rho) = H^T(x_\rho) W H(x_\rho) \tag{2.8}$$

with x_ρ being the estimated x for iteration ρ and the iteration step delta derived as

$$\Delta x_\rho = (G(x_\rho))^{-1} H^T(x_\rho) W (z - h(x_\rho)) \tag{2.9}$$

The iteration state is then updated by

$$x_{\rho+1} = x_\rho + \Delta x_\rho \tag{2.10}$$

The iteration process terminates when the magnitude of the state update falls below a predefined threshold, satisfying the convergence criterion:

$$|x_{\rho+1} - x_\rho| < \varepsilon \tag{2.11}$$

where ε is a small positive value that defines the required precision of the solution. This ensures that further iterations would not significantly improve the state estimate. [15], [26]

Additional Conventional Algorithms

Although the WLS is among the most common conventional state estimation algorithm, it is not the only one. The different algorithms can be divided into static and dynamic SE algorithms [27]. While static SE algorithms determine the steady-state of the system, dynamic SE algorithms aim to capture the dynamics of the system state [28]. Next to the aforementioned WLS, static SE algorithms also include the Least-Absolute-Value algorithm. The Least-Absolute-Value algorithm uses the absolute difference instead of the squared differences as the optimization basis. As the WLS can be susceptible to bigger measurement errors [29], the Least-Absolute-Value algorithm can deliver more robust results [30]. Additionally, Hachtels's-Augmented-Matrix method was proposed to increase numerical stability of the state estimation problem [7], [31]. The method has been applied recently for Low Voltage (LV) distribution grids [32], [33]. A central algorithm for dynamic SE is the Extended Kalman Filter, an extension of the Kalman Filter supporting non-linear equations. The method uses a sequence of historical measurements instead of a state snapshot to estimate the state of a process [34]. The algorithm has shown to increase esti-

mation results compared to WLS in certain scenarios [35]. Recently the authors in [36], [37] have had considerable success with tensor completion methods. In tensor completion, missing or unobserved entries are imputed, utilizing signal sequences.

2.2 Measurement Types and Processing

Effective monitoring and operation of power grids rely heavily on accurate measurement data. This data provides insights into the real-time state of the system and forms the main data source of SE algorithms. This chapter details the common measurement technologies employed in DGs. Furthermore, the concept and application of pseudo-measurements are discussed as a means to augment system observability. Finally, the chapter examines the standards and implications of measurement accuracy.

2.2.1 Measurement Infrastucture

Grid monitoring and grid operation functionality are surfaced in a Supervisory Control and Data Acquisition (SCADA) system. All measurement data from devices in the field are preprocessed and supplied in real-time. While measurement infrastructure in the higher voltage levels is comprehensive, the same can not be said for distribution grids. The technical specifications of available measurements depend on the specific type of measurement device and its technical implementation. Broadly, the common categories of available measurements are those installed in intelligent substations, Micro-Phasor Measurement Units (μPMUs) and Smart Meters.

2.2.2 Intelligent Substations

The digitization in Medium Voltage (MV) distribution grids is mostly driven by the installation of intelligent MV/LV substations. In addition to the MV/LV transformer, they are equipped with measurement infrastructure, communication functionality, and varying degrees of control and protection abilities. Current research focuses on intelligent devices that combine control and protection functionalities based on available decentralized measurements [38]. Measurements in those substations are mostly installed on the LV side of the transformer and the type of measurements that are recorded depend on the specific hardware implementation. Communication is typically implemented via protocol standards IEC 60870-5, IEC 60870-6 and IEC 61850 [5]. Synchronization is achieved by NPT, achieving syn-

chronization accuracy within a few milliseconds [38]. While sufficient for many applications, it is not accurate enough for synchronizing high-frequency signals like voltage angles.

PMUs and μPMUs

Phasor Measurement Units (PMUs) measure voltage and phase angle simultaneously across different locations of the power system while utilizing a shared time source for synchronization. Synchronization across different locations is achieved through precise timing signals provided by the Global Positioning System (GPS) [39], [40]. The synchronicity of measurements allows for a true simultaneous SE [40], [41]. The output of a PMU typically includes the magnitude and phase angle of voltage, frequency, rate of change of frequency, and a timestamp [42].

PMUs are mostly employed in the transmission system as their limitations hinder DG adoption. Firstly, their relatively high cost is unsuitable for broad application in DG [42]. Secondly, due to shorter distances and less current magnitudes in DG, power flows, angle differences are typically two orders of magnitude smaller. Additionally, the signal-to-noise ratio is lower due to generally smaller signal magnitudes. Because of that, the precision requirements of the measurement device are higher [43]. The μPMU has been developed to overcome those shortcomings. Not only do they cost an order of magnitude less, but they also address the specific accuracy demands [43], [44]. Field application of μPMUs encompass a broad range of monitoring and protection functions encompassing

- DSSE
- Topology Detection
- Phase identification
- Event Detection
- Fault Location

among others [43].

Smart Meters

Traditional power meters are being replaced by smart meters. Due to the Smart-Meter-Gateway communication interface, live measurements can be sent to a third-party. Depending on the manufacturer and model, the measurement devices can

measure voltage, current as well as active and reactive power in at least 15-minute time intervals.

One significant difference between PMU-based and smart meter-based DSSE algorithms is the availability of phase angle data. In systems with only smart meters, many researchers have assumed that the nodal voltage phase angles are almost equal, known as the "small phase angle difference" assumption [45]. However, the addition of voltage phase data or flow measurements from PMUs can significantly improve the performance of estimation and identification routines [46].

Different national political schemes and directives have been set to foster the smart meter rollout. Furthermore, rules, and regulations apply to the potential usage of the collected data, which are tied to European data protection laws [47]. In Germany, the transmission of grid state information to the responsible grid operator is currently only allowed under certain condition. Noted use cases include grid state estimation and managing of grid congestions as part of the § 14a [4].

Pseudo-Measurement

Besides direct physical measurement of state variables, pseudo-measurements represent a class of representative values that substitute actual measurements. Given the lack of observability in DGs, their use aims to augment the state information of the system artificially. There are a wide variety of different kinds of pseudo-measurements and many methods to obtain them. Common examples include standard load profiles or model calculations for substituting the power demand or generation of actual systems [48]. Given the inherent inaccuracy of pseudo-measurements and the variability of RES, applying traditional WLS methods has been difficult [49]. In DSSE pilot field applications, a variety of pseudo-measurements are used. The authors of [50] use the maximum historical bus load, current weather and bus generation capacity to distribute the measured load throughout the individual buses. In [51], forecasted load and generation values are considered based on previous measurements. To estimate reactive power values, in [52] a Q(V)-droop controller is used. The authors of [53] use weather nowcasts to infer renewable energy production to use as pseudo-measurements.

2.2.3 Measurement Accuracy

Measurement device accuracy is specified by their associated accuracy class. Typical accuracy classes are 1, 3, and 5, as well as 0.1, 0.2, and 0.5 for more precise

measurements [54]. The numerical value of the accuracy class defines the maximum permissible deviation under reference conditions. For a measurement device with accuracy class 1, the maximum permissible deviation under reference condition would be 1 % concerning the measured value. Requirements for smart meter accuracy have been specified in a corresponding EU directive.[1] The directive defines measurement accuracy classes A, B, and C. These correspond to numerical classes of 3.5, 2, and 0.7 for three-phase loads or 4, 2.5, and 1 for single phase load given a three-phase measuring device.

2.3 Supervised Learning and Neural State Estimation

Artificial Neural Networks (ANNs) are powerful universal function approximators capable of learning complex patterns from data. Their ability to model non-linear relationships across diverse data sets makes them well-suited for a wide range of applications in electrical engineering, including distribution system state estimation. Depending on the task and available data, there are different training modalities, including unsupervised, semi-supervised, and supervised learning. This thesis focuses on supervised learning techniques.

In supervised learning, the model is trained on a labeled dataset of input-output pairs, where the output or labels serve as ground truth. Two fundamental application areas for supervised learning are regression and classification tasks. In regression tasks, the model learns to predict continuous values. For instance, estimating nodal voltages in an electrical grid can be interpreted as a regression task. In classification tasks, the model learns to assign inputs to discrete categories or classes, such as identifying anomalous measurements.

2.3.1 Multilayer Perceptrons

Deep feedforward neural networks, also known as multilayer perceptrons (MLPs), form the foundation of many modern neural network architectures. These networks consist of multiple layers of interconnected neurons, where information flows unidirectionally from the input layer through one or more hidden layers to the output layer.

[1] Directive 2014/32/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of measuring instruments

The mathematical representation of a feedforward neural network can be expressed as:

$$y = \sigma^{(l)}(W^{(l)}\sigma^{(l-1)}(W^{(l-1)} \cdots \sigma^{(1)}(W^{(1)}x + b^{(1)}) + b^{(l-1)}) + b^{(l)}) \tag{2.12}$$

where x is the input vector, y is the output vector, l is the number of layers, W_i and b_i are the weight matrix and bias vector for layer i, respectively, and σ_i is the activation function for layer i. A visual representation of a multilayer perceptron is depicted in Figure 2.2.

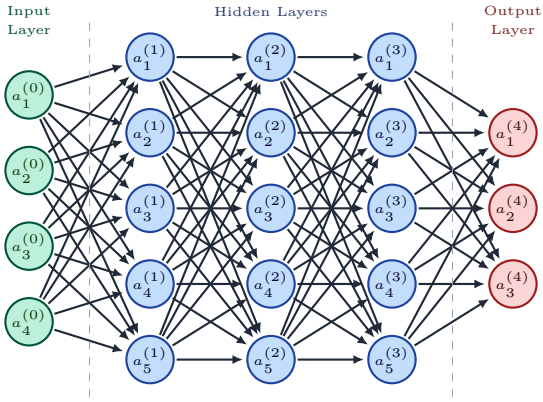


Figure 2.2: Structure of a Multilayer-Perceptron with three hidden layers.

A detailed representation of how to calculate single neuronal output and the resulting layer output is represented in Figure 2.3.

2.3.2 Activation Functions

Activation functions introduce non-linearity into neural networks, enabling them to learn complex patterns and relationships in the data. Without activation functions, neural networks would be limited to learning linear transformations. The choice of activation function significantly impacts the network's learning capability and performance [9].

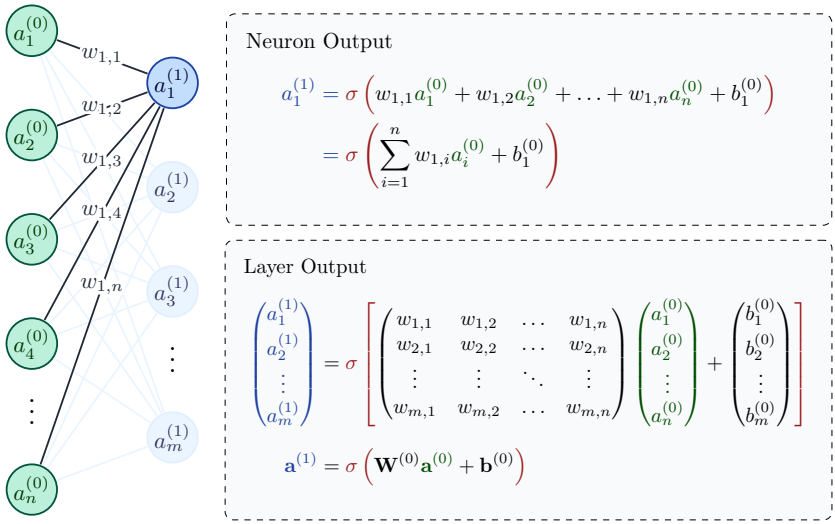


Figure 2.3: Singular neuron and neural network layer output calculation.

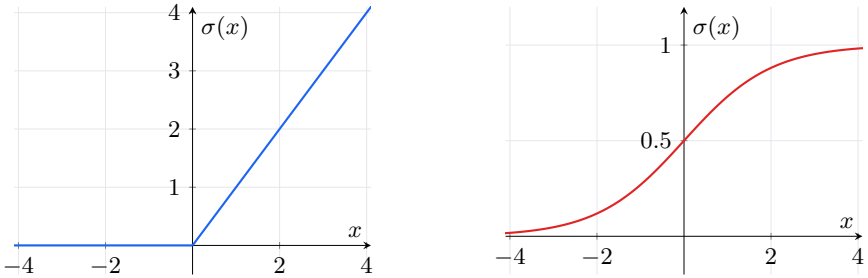
Most commonly rectified linear units are recommended for hidden layers in neural networks [55], [56]. The Rectified Linear Unit (ReLU) function depicted in Figure 2.4a and is computed as follows:

$$\sigma(x) = \max(0, x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \tag{2.13}$$

One practical concern is the so-called "dying ReLUs" phenomenon, where neurons end up outputting exclusively zero during training. To address this issue, several ReLU variants have been introduced such as the leaky ReLU [57], the exponential linear unit (ELU) [58] and the scaled exponential linear unit (SELU) [59].

The activation function used in the output layer depends on the task. For regression tasks, the identity function is commonly used. Classification tasks mostly use the maximum likelihood method given a calculated set of probabilities for different classes. The sigmoid function is a common choice to compute probabilities in the output layer [9]. It is depicted in Figure 2.4b and calculated as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.14}$$



a) ReLU Function

b) Sigmoid Function

Figure 2.4: Visualization of ReLU and Sigmoid activation functions.

2.3.3 Training Neural Networks

For ANNs to calculate anything useful, they have to be trained. The following description is based on [60]. Neural network training is the process of adjusting the ANN parameters, the weights and biases, to minimize a predefined loss function $\mathcal{L}$. The loss function quantifies the models' performance on the given input data. A common loss function is the mean squared error:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.15}$$

where y_i is the true and $\hat{y}_i$ the predicted value for training sample i.

Given a quantified loss, the training process is an optimization procedure that minimizes the loss by updating the trainable parameters. The most common optimization method is the iterative gradient descent method. For each iteration, the partial derivative of the loss function for each trainable parameter is calculated. The resulting gradient of the loss function represents the direction that reduces the loss function most rapidly. Each parameter is updated in gradient direction with

a predefined step size called the learning rate. Given a loss function $\mathcal{L}(\theta)$, where θ represents all trainable parameters of the network, the gradient descent update rule is:

$$\theta_{\rho+1} = \theta_\rho - \eta \nabla_\theta \mathcal{L}(\theta_\rho) \tag{2.16}$$

where η is the learning rate that controls the step size, and $\nabla_\theta \mathcal{L}(\theta_\rho)$ is the gradient of the loss function with respect to the parameters at iteration ρ.

The computation of gradients over the entire training dataset is computationally expensive. Instead, Mini-Batch Gradient Descent and Stochastic Gradient Descent compute gradients on small random subsets of the data called mini-batches or singular training samples, respectively. In this way, they represent a trade-off between computational efficiency and gradient accuracy.

2.3.4 Feature Scaling

ANN training features can have significantly different scales. While voltage magnitude values, if noted in p.u., typically range from 0.9 to 1.1, the range of power values is much wider. Due to the gradient-based optimization this can destabilize training and reduce convergence speed. Feature scaling is a preprocessing step that harmonizes the feature scales which alleviates that concern. Common feature scaling techniques include:

- **Min-Max Scaling:** Scales features to a fixed range of typically [0, 1]:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2.17}$$

- **Standardization:** Transforms features to have zero mean and unit variance:

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma} \tag{2.18}$$

  where μ is the mean and σ is the standard deviation of the feature.

Scaling parameters, like $x_{\min/\max}$, μ and σ, should be computed only on the training data and then applied to validation and test sets to prevent data leakage [60].

2.3.5 Hyperparameters and Hyperparameter Tuning

Unlike the trainable parameters (weights and biases) hyperparameters are set before training and control specifics like model size and the training process. Choosing a proper set of hyperparameters significantly impacts training and model performance. Some of the key hyperparameters, as mentioned in [9], include:

- **Network Architecture:** The amount of hidden layers and neurons per layer. Additional layers and neurons increases the model's complexity but can lead to overfitting.
- **Activation Function:** The non-linear function that gets applied to neuron outputs.
- **Weight Initialization:** Method to set initial weights before training. Differences in weight initializations can influence convergence behavior.
- **Learning Rate:** The step size during gradient descent optimization. While too large values can cause unstable training, too small values hinder the convergence of the model.
- **Number of Epochs:** The number of complete passes through the training dataset during training.
- **Batch Size:** Controls the number of training samples in each batch during Batch Gradient Descent. Larger batches can provide more stable gradients but require more memory and might increase the overall training duration.
- **Optimizer Configuration:** Algorithm specific parameters that control the optimization process.
- **Regularization Parameters:** Parameter settings to control overfitting prevention, such as dropout rate or L1/L2 regularization strength.

Searching for the best combination of hyperparameters is called hyperparameter tuning. Since evaluating a set of hyperparameters requires an entire model training run, this search can become prohibitively computationally expensive. Some of the standard procedures are listed in [61], [62] and include:

- **Grid Search:** Evaluates all combinations from predefined sets of hyperparameter values. While comprehensive, this becomes computationally expensive with many hyperparameters.
- **Random Search:** Samples hyperparameter combinations randomly from predefined ranges. It can be more efficient than grid search, particularly when hyperparameters have varying levels of importance.

– **Bayesian Optimization:** Uses probabilistic models to guide the search for optimal hyperparameters. The optimizer learns from previous evaluations to focus on promising regions of the hyperparameter space.

2.3.6 Neural Networks for State Estimation

Early applications of ANN for SE date back to 1991 in a transmission system context [63]. The rise in distribution system complexity and advanced Machine Learning (ML) methods have lead to increased research in the field in recent years [64]. State-of-the-art DSSE methods utilizing neural networks can be categorized into three primary approaches:

Purely data-driven approaches

These methods rely solely on neural networks to estimate the grid state from the available measurements. The foundational architecture is the deep feed-forward neural network. This model has been successfully trained and utilized given labeled synthetic data [65]. In [66], Autoencoders were used as a basis for DSSE, utilizing two separate ANN for encoding and decoding measurement state vectors. An Attention-Enhanced Recurrent Neural Network for modeling pseudo-measurements was proposed in [67]. This method utilizes power series analysis at the source and load in the time and frequency domains. It creates a period-dependent extrapolation model that characterizes the power series in these domains.

Physics-Informed Neural Networks

These methods integrate the physical laws and topology of the power grid into the architecture of the neural network. The authors in [13] integrate the grid parameters into a physics-informed loss function for the neural network. The inclusion of the grid topology can reduce the number of coefficients needed to parameterize the mapping from measurements to the grid state. For example, a Physics-Aware Neural Network (PAWNN) was proposed in [68], exploiting the structure of the power grid to reduce the number of coefficients required to realize the state estimation mapping. It does so by exploiting the separability of the estimation problem, thus preventing overfitting and reducing the complexity of the training stage. Unrolling an iterative physics-based prox-linear solver was proposed in [69] as a physics-based Neural State Estimation (NSE) approach. The methods outlined in [68], [69] and

additional graph-based architectures from [70], [71] were compared in [P3]. It was found that recurrent graph convolutional models performed comparably to more specialized models at lower model complexity. GNN-based state estimators have also been proposed in [12] that model the state estimation problem in distribution networks as node-level prediction problems on their graph representations with state measurement matrices and tensors as input features. A Graph Convolutional Network based approach for state estimation was presented in [72] that directly integrates the graph structure of the network into the filter matrix of the GCN.

Hybrid approaches

These methods combine neural networks with traditional model-based state estimators. For instance, neural networks can generate pseudo-measurements or pre-initialize unknown measurements for conventional state estimators [73], [74]. A hybrid state estimator is presented in [73], which uses an ANN for topology identification and a set of ANNs combined with a WLAV algorithm for state estimation. This hybrid approach combines the strengths of data-driven and model-based approaches to ensure robust estimation results given different time resolutions of measurements. Given incomplete measurement data, a trained ANN estimates during limited observability states. A weighted least absolute value approach is used in case of full observability of the network. The authors in [74] use a ANN to create an initial state estimation that gets refined using conventional optimization methods. In [75], neural networks estimate the bus injections from real-time measurements. These estimated bus injections can then be used as pseudo-measurements to compensate for the lack of real-time measurements.

2.4 Distribution System State Estimation Challenges

Implementing DSSE presents unique challenges distinguishing it from traditional state estimation approaches used in transmission systems. These challenges arise from the distinct characteristics of DGs and their environmental properties. Several technical, operational, and implementation challenges must be addressed for successful deployment of the DSSE. This section systematically examines these challenges from two perspectives: general challenges inherent to DSSE and specific challenges that emerge when applying supervised learning techniques to this domain. Understanding these challenges is essential for developing robust and practical DSSE solutions that can meet the particular requirements.

2.4.1 General Challenges

Distribution System Observability

The central challenge the DSSE has to handle is the DG's limited observability due to limited measuring infrastructure. Different advanced algorithms and estimation techniques are developed to overcome this hurdle. While WLS-based approaches tend to utilize pseudo-measurements as measurement replacements, other types of algorithms and computational methods, such as ANNs, do not require full grid observability [46].

Grid Topology Identification

One of the most influential factors on the voltage of individual nodes is their relative position in the graph model. The topology identification problem can be classified as a system configuration or a topology learning problem. Firstly, the system configuration identification problem deals with identifying the current operating topology of the grid. The basic topology is known, but local events like switching, faults, and line disconnections might not be registered [46]. Possible solutions include the analysis of voltage statistics [76], [77], a recursive Bayesian approach [78], [79], using an Autoencoder [80], among others [81], [82]. Secondly, the topology learning problem deals with the fact that the system operator might not have complete or even partial knowledge of the basic grid topology. Given the bus and branch measurements, the objective is the discovery of the underlying graph topology. [46] Using a sparse graph recovery algorithm, authors in [83] leverage the DC power flow, while the authors in [45] rely on the probabilistic relationships between different voltages.

Unbalanced Grid Operation

The transmission system is mostly symmetric, meaning there is little difference in state between the phases other than their phase shift. In such cases, the SE problem can be simplified to a single-phase SE [7]. The situation can be different in DGs, particularly in low-voltage grids. One driver for increases in phase-imbalances is the rise in RES. Since asymmetrical infeed to a maximal imbalance of 4.6 kVA is allowed according to the applicable rule [84], a significant amount of Photovoltaic (PV) plants are connected to a single phase [85]. Therefore, a three-phase estimation model might be necessary to achieve satisfactory precision [15]. The authors in [86] have investigated a three-phase low-voltage state estimation using the conventional

WLS approach. When using ANNs for state estimation, the input of the models can be extended to incorporate all individual three-phase measurements and output results for all three phases accordingly. Alternatively, the model can be fed with the input combination for the three phases individually [65].

Differing Measurement Time Resolutions

Given the diversity of measurement devices and their different technical implementation, measurement signals might have different resolutions. That leads to differing measurement vectors that complicate the SE computation. To account for the different measurement combinations, the authors in [87] introduce a dynamic state estimation. Each estimation is a function of the previous estimate, the variable amount of gathered measurements, and inertia parameters. This approach is compared to a WLS approach in [88]. Here, additional measurements are considered pseudo-measurements with a finite time window.

Asynchronous Measurements

While PMUs and μPMUs are usually synchronized using a GPS signal, the same is not true for smart meters. The resulting asynchronicity can lead to worse estimation results [8]. In [89], the authors synchronize the measurements by maximizing the temporal correlation between the different data signals.

Measurement Inaccuracies, Faults, and Missing Signals

Measurements contain some level of measurement inaccuracy (see Subsection 2.2.3). The SE needs to be sufficiently robust to deal with the noise level. Furthermore, measurement faults can occur due to device failure or failure of the communication infrastructure, leading to faulty or missing data. Research at the intersection of power systems and ICT investigates SE performance with respect to combined power system and ICT requirements [90]. One solution to handle grave errors or missing signals is to reconstruct such signals. In [80], the usage of Autoencoders for signal reconstruction is proposed.

Cybersecurity Concerns

As more and more measurement and communication infrastructure is integrated into the DSSE, and increasing monitoring and operational measures are introduced,

the surface area for potential cyberattacks and, therefore, the associated risk increases. Such attacks aim mainly to either manipulate data collection or disrupt data communication [91]. Current research investigates the establishment of trust values for physical measurements that capture not only possible cyberattacks but also other sources of measurement errors and inaccuracies [92]. In [93], the authors leverage the trustworthiness of input measurements to estimate the trustworthiness of the SE state variables.

Low X/R Ratio

A common way to simplify the state estimation problem and improve computational speed relies on the assumption of a low resistance-to-reactance ratio of transmissions system lines. DG lines exhibit higher resistance [25]. Since this invalidates the basic assumption of the simplification, these methods have become infeasible [94].

2.4.2 Supervised Learning Challenges

A supervised learning system uses labeled training data to learn a task given appropriate model training. This specific modality raises certain practical challenges and poses specific questions about their reliability that are not discussed much in current research.

Training Data Generation

Most fundamentally, supervised learning techniques learn from labeled training data. Ideally, this would be data from the actual system. Given the nature of the issue the DSSE is trying to solve, the electrical grid for which the DSSE should be utilized is only sparsely measured. In effect, no complete data set for training is available, and different means of generating sufficient training data are necessary. Most literature approaches solely rely on generated synthetic data. This approach, however, does not apply to real-world systems where the DSSE would be employed. Actual training data generation is, therefore, more complex [72].

Changes of Operation State

When training a supervised model, that model gets fitted to the particular underlying data distribution of the training data. Even if the training data is a realistic

representation of the measurement distribution initially, the measurement distribution might change over time. This could be due to new consumers and generators being added or removed from the grid. If enough of those changes occur, the training and measurement distribution drift apart, leading to worse estimation results also known as data drift [95]. Firstly, the sensitivity of trained models to those changes in generation and consumption patterns should be analyzed. Secondly, a mechanism that detects the mentioned changes and provides an appropriate response should be in place.

Estimation Confidence

A common critique of ANN and other ML models is their black-box nature. Since these models are not rule-based but are fitted to the data, it is hard to introspect any given estimation. As the DSSE is a monitoring and operation tool for critical infrastructure, trust concerning algorithm performance plays a central role. Thus, any mechanism for increasing confidence in the model estimation or quantification of the estimation error increases trustworthiness and facilitates potential field implementation of the algorithm. Approaches in literature include using bayesian deep unfolding [96] and a probabilistic approach, estimating states with confidence intervals [97].

Model Estimation Validation

Since the DSSE model outputs estimation for unmeasured buses, the model's output lacks a means of result validation. Validation is possible when given a synthetic dataset where the complete system state is known. When using an actual grid model with limited observability, any estimation for unmeasured buses can not be validated by comparison to a ground truth or measured value.

This thesis mainly focuses on the real-world applicability of supervised learning for DSSE. As such, the focus lies, although not exclusively, on addressing the mentioned specific supervised learning challenges. Figure 2.5 depicts an overview of where and how the individual challenges are addressed.

2.5 Summary

This chapter establishes the fundamental concepts and theoretical background necessary for understanding DSSE and its implementation using neural networks. It

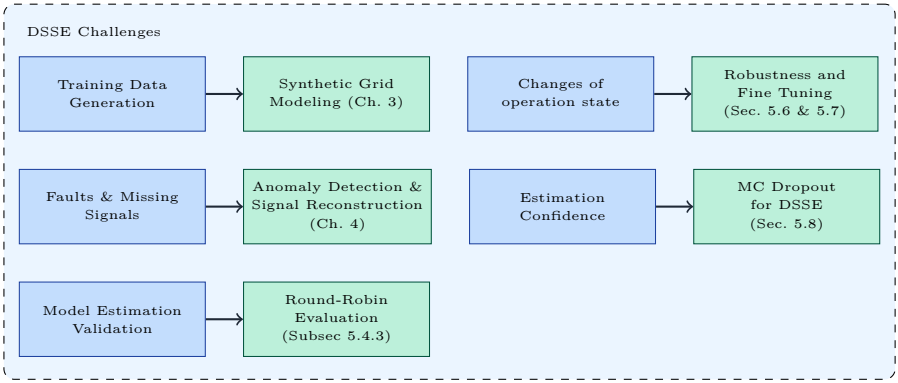


Figure 2.5: Overview of different supervised learning specific DSSE challenges including how and where they are covered in this thesis.

begins by introducing the principles of grid monitoring in distribution systems, explaining how the DSSE serves as an operational monitoring tool for establishing grid observability and managing congestions. The conventional state estimation problem formulation is presented, including the mathematical foundations of the Weighted Least Squares method and other traditional approaches.

The discussion of measurement types and processing highlights the various data sources available for state estimation, from PMUs and smart meters to pseudo-measurements, along with their respective characteristics and limitations. Additionally, quantification of measurement accuracy due to specification of measurement accuracy classes is discussed. The chapter then introduces the principles of supervised learning and neural networks, covering essential concepts such as network architectures, activation functions, and training procedures. The application of neural networks to state estimation is explored, presenting different approaches ranging from purely data-driven methods to physics-informed and hybrid solutions.

Finally, the chapter concludes with a comprehensive examination of both general DSSE challenges and specific issues related to supervised learning applications. These challenges and various practical implementation concerns, especially those related to supervised learning based DSSE form the foundation for understanding the research problems addressed in subsequent chapters. Building on this foundation, the following chapters aim to develop and analyze solutions that bridge the gap between theoretical approaches and practical implementation in real-world DGs.

3 Grid Model and Training Data Generation

In the literature, supervised learning approaches for DSSE are trained, evaluated, and tested solely on synthetic data. The fundamental problem is that the model training requires a complete training data set, which is only present for fully observable grids. Since the problem the DSSE is trying to solve is estimating the state of a not fully observed grid, a complete training data set is not available by definition. Given the rising importance of synthetic grid models, exemplified by the ongoing research into digital twins [98], a potential solution is to utilize synthetic models to simulate the grid's behavior. If the simulation is sufficiently accurate, the simulated data set can be used for model training and transferred to field application.

To explore this approach, this chapter describes the grid modeling and simulation for an existing MV grid. First, the real-world MV distribution grid is described in Section 3.1. Section 3.2 describes the construction of a synthetic training data set, based on that grid model. Finally, the synthetic data set is validated by evaluating the simulation results with respect to the physical field measurements.

3.1 Grid Model Description

As part of the grid model description, the physical system and the measured system state is analyzed. Given the limited digitalization seen in today's distribution grids, one central challenge is data quality. The available data is often extracted manually from documents which is susceptible to human error. Furthermore, it can't be assumed that all of the information is up to date. Since a lot of the information given by the grid operator cannot be verified, one can expect certain inaccuracies. This is true for the electrical grid model but especially for supplemental details on connected systems. The data set used here is no exception. While this makes some evaluations more challenging, it is a realistic representation of implementing this methodology in the field.

3.1.1 Electrical Grid Model

The electrical grid used to validate the real-world performance of the developed models is a 20 kV MV grid. The grid contains two branches. Only one of those two branches is equipped with measurement devices. Since the unobservable branch does not allow any validation, the model was reduced to the measured branch. This

was possible since there is no electrical connection between both branches, making them electrically independent.

The reduced model consists of 48 nodes and 46 lines with an aggregated line length of about 30 km. A graph model of the grid is depicted in Figure 3.1. While the graph model presents the general graph structure, the graph was morphed to ensure data protection. Of the overall 48 nodes, 19 nodes are equipped with measurement devices.

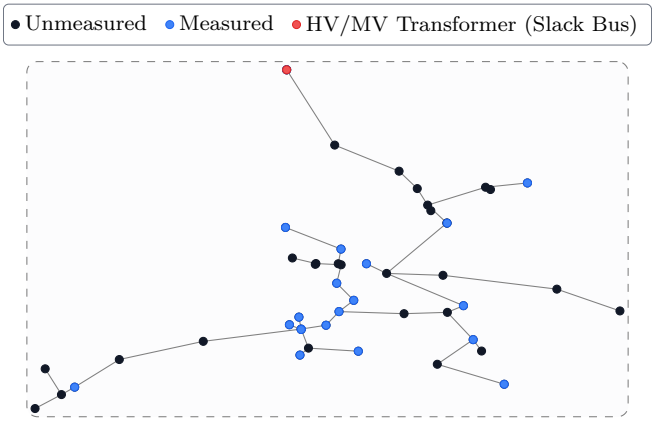


Figure 3.1: Schematic grid model diagram of the MV grid's measured branch.

3.1.2 Measurement Devices

The installed measurement devices installed record the following values:

- Three-phase active power
- Line-to-ground voltage
- Voltage angle
- Active Power

Except for the three-phase active power, all values are recorded for all three phases individually. All measurements are installed at the LV side of the MV/LV trans-

former. The measurement devices have a specified accuracy class of 1, which means the maximum error is expected to be 1 % of the measured value (see 2.2.3). Unfortunately, the angle measurements are not precisely gps-synchronized. Given its 50 Hz frequency, minor asynchronies make it unusable for validating angle estimations in the SE.

3.1.3 Connected System Participants

Table 3.1 lists registered systems that are connected to the grid. The supplied area is primarily residential. Therefore, household appliances and their respective heating demand incur most of the load. Concerning distributed generation, a high share of PV systems can be observed.

The quantities in Table 3.1 have some associated uncertainty. The numbers themselves might not be up to date, and several unregistered systems are expected. Furthermore, no precise information about individual systems and their parameters is known.

Table 3.1: Overview of registered systems within grid region

System	**Count**	**Power [kW]**
Household	671	-
PV plant	219	3277
Night storage heating	112	1718
Direct heating	77	237
Heat pump	30	116
Storage	7	38
EV charging station	7	118
Other	87	270

3.1.4 Measurement Preprocessing

One of the major differences between synthetic and real data is the existence of faulty measurements and measurement inaccuracies. Furthermore, voltage scaling is necessary for comparing measured voltages due to varying transformation ratios

of the transformers. This section outlines the corresponding preprocessing steps to prepare the measurements for this thesis.

Voltage Scaling

Voltages in the context of the DSSE are mostly presented in p.u. units with respect to the rated voltage of the grid level. Since the measurement devices measure absolute values on the LV side of the MV/LV transformer the measurements are scaled accordingly. To relate the LV side measurement to the MV side voltage, the voltage is scaled using the voltage ratio of the respective transformers, taking into account their set tap positions. If we consider $V_{i,rated}^{A(B)}$ to be the nominal voltage of the higher (lower) voltage side of the transformer i, λ_i its' tap position and dV_i the relative voltage magnitude increase per tap position, one can calculate the voltage transformation ratio Λ_i as follows:

$$\Lambda_i = \frac{V_{i,rated}^{A}}{V_{i,rated}^{B}} \times (1 + \lambda_i \times dV_i) \tag{3.1}$$

The voltage measurements $v_i(t)$ in p.u. can then be calculated as:

$$v_i(t) = V_i(t) \times \Lambda_i \tag{3.2}$$

where $V_i(t)$ is the LV side voltage magnitude measurement. To get the precise MV-side voltages, the respective voltage drops of the transformer would need to be considered. As no technical specifications of the MV/LV transformers were available, these were neglected.

Faulty Measurements

Measurements from field devices are subject to numerous potential influences that can lead to wrong or missing signals. Potential sources of errors include unregistered tap changes, signal loss, or technical errors of either the measurement device or the communication infrastructure. Figure 3.2 shows two examples of potentially faulty measurements from the data set. In both examples, the potentially faulty measurement was contrasted alongside the electrically closest measurements. The electrically closest measurements are those with least aggregated impedance between the individual buses. Subplot a) clearly shows that a lost or faulty signal leads to constant measurements throughout the day. Subplot b) becomes conspicuous when

compared to the closest measurements. Given the tight correlation of close measurements, a sudden jump in their relative differences is unlikely. The jump in voltage is best explained by a unregistered tap change of the corresponding transformer.

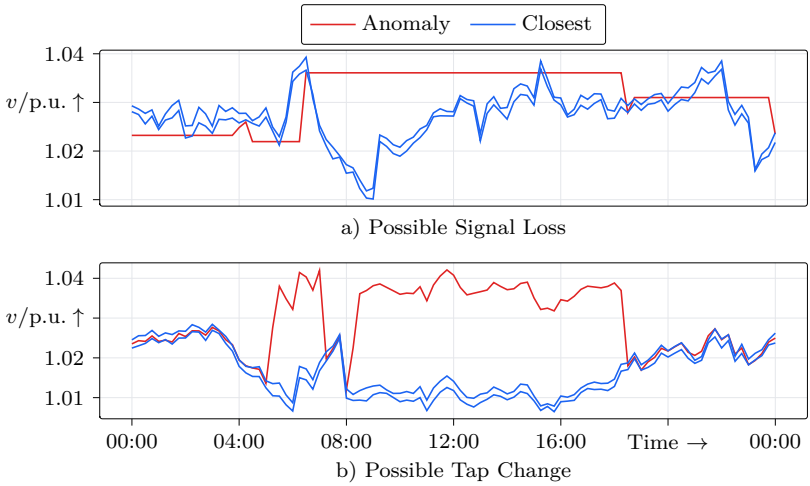


Figure 3.2: Exemplary suspected anomalies of the measurement data. "Closest" denotes the electrically closest measurements to the suspected measurement anomaly.

Filtered Measurements

As the measurements will be used to assess model performance throughout this thesis, faulty measurements pose an evaluation problem. When the output of a model is compared to a measurement, a faulty measurement as the expected value will lead to incorrect assumptions about the model's accuracy. For that reason, a filtered data set was created. As referenced in Figure 3.2, while the signal loss can be identified due to a constant signal, identifying other measurement faults is more complex and can not be easily verified. Consequently, only constant signals were extracted from the measurement data. A measurement is considered constant if its

values do not change for at least three consecutive time steps. Chapter 4 deals with detecting other anomalies.

Measurement Noise

Given the accuracy class of 1 of the measurement devices, raw measurements are expected to have a maximum error of 1 % (see 2.2.3). Random measurement noise is expected to be approximately normally distributed. If one assumes the maximum error $e_x^{\%}$ of the true value x to not be exceeded within three standard distributions, the standard deviation of the measurement σ_x of the random noise can be calculated by:

$$\sigma_x = \frac{x \times e_x^{\%}}{3 * 100} \tag{3.3}$$

A detailed explanation of the formula can be referenced in [99]. The same or a similar procedure of modeling measurement error has been used in similar publications [22], [65]. Figure 3.3 depicts a synthetic voltage magnitude signal with noise, sampled from a normal distribution, given the noise standard deviation as outlined in Equation 3.3 for a measurement class of 1.

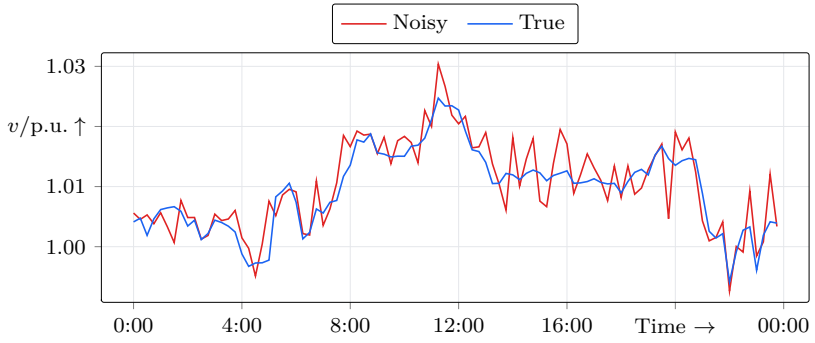


Figure 3.3: Synthetic measurement signal with generated measurement noise.

Figure 3.4a) depicts measurements of three nodes located directly next to each other. One would expect the underlying true voltage to be very similar, so their relative

difference indicates the level of measurement noise. Subplot b) shows a measured signal with synthetically applied noise. Each time series has the same underlying signal but independently sampled noise based on Equation 3.3. Given the specified measurement accuracy, this represents the expected relative difference of similar signals. The expected signal difference due to noise and the actual measurement similarity of close measurements suggests that some measurement preprocessing or averaging was applied. Unfortunately, no specific information thereof is known, which complicates later analyses.

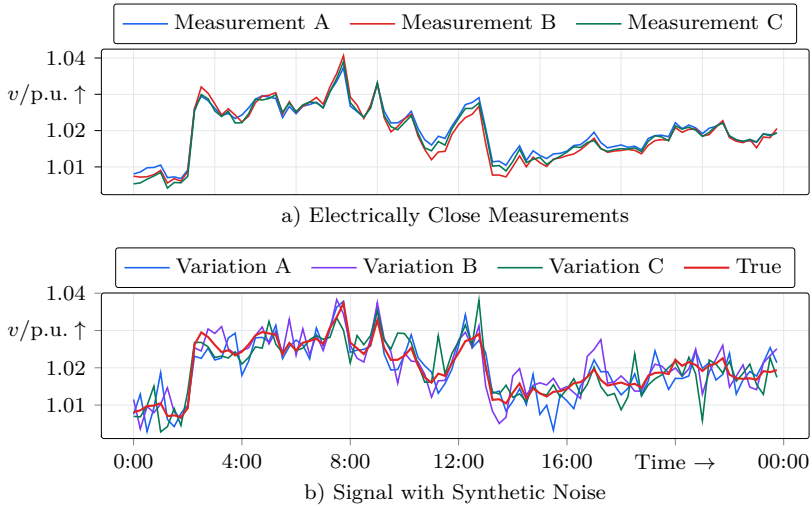


Figure 3.4: Close measurements with actual noise and corresponding signal with synthetically generated noise.

3.1.5 Measurement Evaluation

A base analysis of the measurements was performed to understand the overall grid behavior. The analyzed values are measured throughout the year 2023.

Table 3.2: Minimum, mean, and maximum spread statistics of measured bus voltages

	Minimum in pu	**Mean** in pu	**Maximum** in pu
Minima	0.9967	1.0013	1.0089
Means	1.0197	1.0240	1.0270
Maxima	1.0468	1.0642	1.0754

Voltage Evaluation

The bus voltage distribution is shown in Figure 3.5. Voltages rarely fall below 1.0 p.u. and do not exceed 1.08 p.u.. Given the formal voltage range of 0.9 to 1.1 p.u. no formal voltage congestions are measured. Furthermore, 90 % of all measured voltages fall into the range of 1.01 to 1.04 p.u.. Minimum, mean and maximum statistics are notated in Table 3.2. Each row represents the minima, mean and maxima of all bus voltages and the column the mean, minimum and maximum values of these respectively.

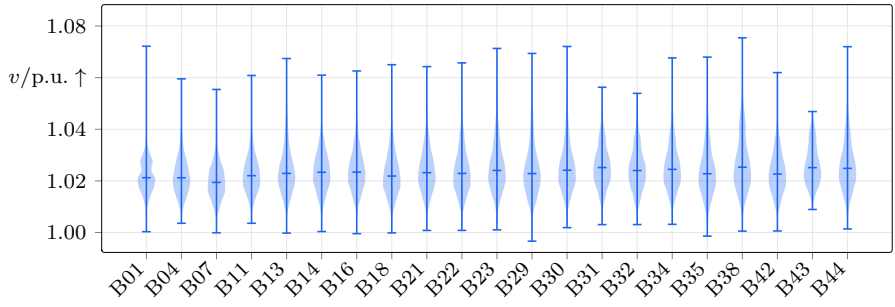
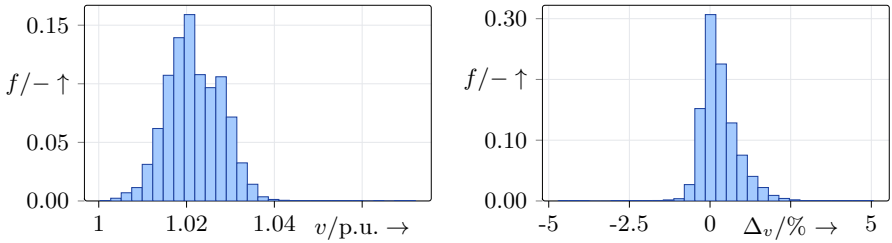


Figure 3.5: Measured MV side bus voltage distributions.

Figure 3.6a depicts a histogram of the HV/MV transformer voltage magnitude. To calculate the voltage spread within the MV grid, the deviation between the individual bus voltages and the HV/MV transformer voltage was calculated. The

voltage deviation histogram is shown in Figure 3.6b where f denotes the normalized frequency. Most of the time, the voltage deviation is relatively small, with 90 % of deviations between -0.40 and 1.36 %. The deviation is skewed positively indicating, that higher deviations are caused by infeed rather than load. The long tails of the histogram show that rare outlier events occur with deviations of up to -5 and 5 %.



(a) HV/MV Voltage Distribution

(b) MV Voltage Deviation

Figure 3.6: Relative frequencies for the MV-side voltage magnitude of the HV/MV connection and the voltage deviations of all MV nodes with respct to the MV-side HV/MV connection.

Power Evaluation

The measured bus power distribution is visualized in Figure 3.7. Positive values indicate load and negative values net generation. Minimum, mean and maximum statistics are notated in Table 3.3. The amount of load and generation differs widely between the respective nodes. Most notably, Bus 14 contributes the highest load and infeed. While load on average tends to be positive, most buses have considerable PV generation resulting in high generation peaks, that oftentimes surpass the load peaks. This indicates the challenge for distribution grids caused by high simultaneity of renewable generation.

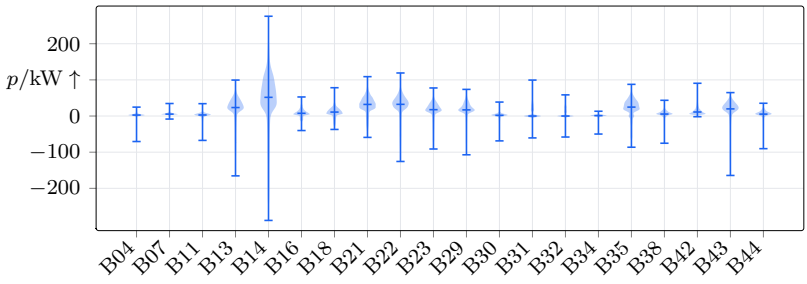


Figure 3.7: MV side bus active power distribution for all measured buses.

Table 3.3: Minimum, mean, and maximum spread statistics of measured bus power

	Minimum in kW	**Mean** in kW	**Maximum** in kW
Minima	-323.80	-97.14	-1.91
Means	-26.82	7.12	34.66
Maxima	13.26	75.01	275.98

Measurement Asymmetries

To investigate the measurement asymmetry the individual phase voltage deviation from the mean voltage is calculated. The mean voltage is calculated as follows

$$\bar{v}(t) = \frac{v^a(t) + v^b(t) + v^c(t)}{3} \tag{3.4}$$

Voltage deviations here and for the remainder of this thesis are given in percentage deviation from 1 p.u.. This is chosen since voltage deviations are small overall. Hence, deviations of 0.001 p.u. are harder to read than 0.1 %. The phase voltage deviations for phase $\gamma \in \{a, b, c\}$, $\Delta v^\gamma(t)$ follow:

$$\Delta v^{\gamma}(t) = (v^{\gamma}(t) - \bar{v}(t)) \times 100\ \% \tag{3.5}$$

and are shown in Figure 3.8.

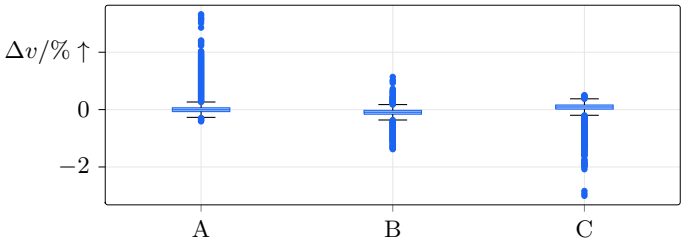


Figure 3.8: Phase voltage deviation from mean voltage in %.

The IQR of all three plots is below 0.15 %, with the most extreme outliers ranging from −3.00 % to 3.32 %. In total only 1.5 % of values exceed a deviation of 1 % or −1 %. Some of those might be due to measurement errors rather than electrical asymmetry. Given that the measurements are mostly symmetric a single-phase SE is deemed sufficient for the context of this thesis.

3.2 Synthetic Grid Model

To create a synthetic data set that represents the usage patterns of the real grid, a synthetic grid model is created. The model consists of the specific grid infrastructure, e.g. buses, lines and transformers, as well as the load and generation systems connected to the grid. Using the energy system simulator SIMONA, a time-series simulation of grid usage is performed. To assess the accuracy of the synthetic data set, it is compared against the real-world measurements.

3.2.1 Grid Data Conversion

The grid model is converted via the openly accessible powerFactory2psdm[2], developed to convert the available PowerFactory grid model to the PowerSystemDatamodel

[2] powerfactory2psdm on Github - https://github.com/ie3-institute/powerFactory2psdm

(PSDM) data format. The PSDM is the data format used to specify the grid model and system participants for SIMONA [100].

3.2.2 System Participant Conversion

Grid usage is driven by the grid's load and generation systems, also called system participants. While the conversion of the grid structure is conceptually straightforward, this does not hold for the conversion of the system participants. While the electrical infrastructure is managed by the grid operators, generation and consumption are not. Depending on the simulation environment, a detailed set of parameters to configure system behavior is necessary. Since most of the time, not all physical parameters needed to set the respective simulation model are available, those must be modeled generically.

As noted in Table 3.1, almost all load and generation in this particular grid stems from households, PV plants, and night storage heating systems. For that reason, those systems are focused on for modeling. No information about the LV grids was available. Furthermore modeling each LV grid explicitly would make the grid simulation very expensive. For that reason, all participants of the synthetic grid model are connected directly to the MV nodes.

Household Load

As the basis for modeling households, the number of households was equated to the number of metering points. The average yearly household electricity demand in 2021 was 3383 kWh. The values ranged from 2105 kWh for a one-person household to 5411 kWh for a household with at least three inhabitants [101]. For loads of the synthetic grid model, values were sampled from a normal distribution with a mean of 3500 kWh and a standard deviation of 1000. The values slightly exceed the average to err on the side of more load, given the positive trend of electricity consumption. The chosen standard deviation represents the difference between different household sizes. Each household demand is modeled stochastically within the simulation. The power values are drawn from a parameterized distribution. As shown in Figure 3.9, individual households generate stochastic states for each time step, but the aggregated load behavior resembles the household standard load profile. This has the advantage of sampling realistic but random power values to generate more distinct grid states. The specific implementation details are described in [102].

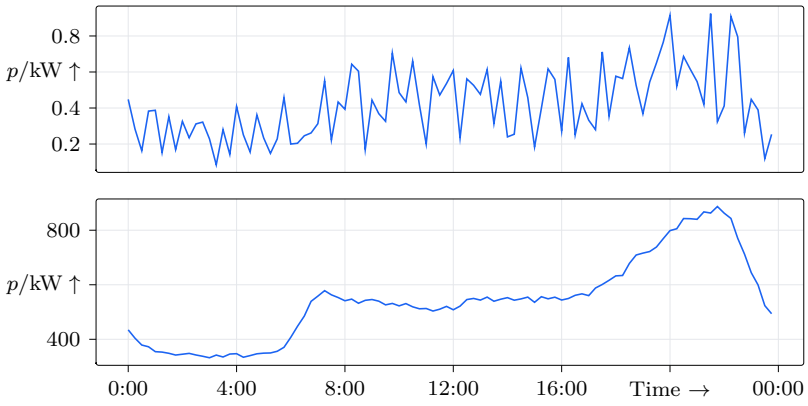


Figure 3.9: Shows singular stochastic household load in the upper plot and the aggregated household load in the lower plot.

Photovoltaics

Regarding PV plants, information about the aggregated amount of systems and the aggregated PV power for each bus was available. Each system was modeled using the mean power. The orientation and elevation angle of the PV systems were sampled using the empirical distribution described in [103]. Around 95 % of all systems had an orientation of either south, east/west, southwest, or southeast. Their relative share is depicted in Figure 3.10. The mean elevation angle of the PV systems is around 32° with 75 % of values within the range from 20° to 44°. When assuming a normal distribution, the standard deviation can be calculated. A z-score table lookup for a probability of 12.25 % results in a z-score of approximately 1.15. The formula to calculate a z-score for a value x, given the standard deviation σ is

$$z = \frac{x - \mu}{\sigma} \tag{3.6}$$

Using the upper interval boundary, one can solve for the standard deviation

$$\sigma = \frac{44 - 32}{1.15} \approx 10.43 \tag{3.7}$$

The elevation angles for the PV systems are sampled from the resulting standard deviation with a mean of 32 and a standard deviation of 10.43.

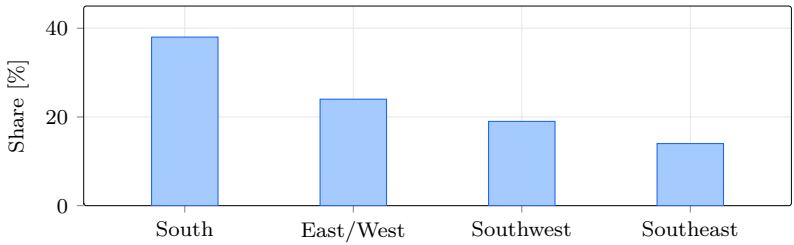


Figure 3.10: Empirical relative share of the orientation of pv systems as described in [103].

Night Storage Heating

Given each bus's installed night storage heating capacity, the corresponding load was modeled by leveraging the published temperature-dependent load profiles. [104]. Temperature-dependent load profiles estimate the load of temperature-dependent systems like night storage heating systems or heat pumps based on ambient temperature. Further information and guidelines for their application can be found in [105]. Each grid provider publishes active power profiles $\tilde{p}^{\mathrm{nsh}}(\theta)$ of active power consumption scaled to 1000 kWh per year, where θ is the average ambient temperature. To relate load profiles to specific devices, they are scaled according to the yearly energy demand of the system. As no information about electricity demand was available for the investigated grid section, the aggregated rated power of night storage heating at the individual buses of the grid was utilized. 100 % of rated power $P^{\mathrm{nsh}}_{\mathrm{rated}}$ was assumed to be drawn during the time interval of maximum load of profile $\tilde{p}^{\mathrm{nsh}}(\theta_{\mathrm{min}})$, where θ_{min} represents the coldest average temperature of the year. Given the information the scaling factor for the Night Storage Heating (NSH) was calculated as follows:

$$s^{\mathrm{nsh}} = P^{\mathrm{nsh}}_{\mathrm{rated}} / \max(\tilde{p}^{\mathrm{nsh}}(\theta_{\mathrm{min}})) \tag{3.8}$$

The NSH profile is then defined as:

$$p^{\mathrm{nsh}}(\theta) = s^{\mathrm{nsh}} \times \tilde{p}^{\mathrm{nsh}}(\theta) \tag{3.9}$$

3.3 Simulation and Validation

Utilizing the converted grid model and system participants model from Section 3.2, this chapter describes the power system simulation and validates the results against field measurements from the real-world grid.

3.3.1 Power System Simulation using SIMONA

A representative usage profile of the grid infrastructure is necessary for subsequent training of the ML models. The grid usage profile is the result of power draw and generation from the grid's system participants. Thus, the participant's behavior over a time interval needs to be simulated and the resulting power flows have to be calculated. This is achieved by leveraging the bottom-up agent-based power system simulator SIMONA [100]. An overview is depicted in 3.11. The simulation model of simona comprises the physical grid topology and the physical system participants models. Each physical system participant model is encapsulated by an agent within the simulation context. The individual agents in turn simulate their behavior based on the underlying physical model. The simplified process for a PV system is depicted in Figure 3.12. During the simulation, each system participant gathers necessary information from environment services, like weather from a dedicated weather service, and then leverages its physical model to calculate its current power output. The net power for each bus is calculated and a power flow analysis determines the resulting bus voltages and line currents. Consequently, the resulting data set consists of complex voltages of all buses, the line and transformer currents as well as the active and reactive power generation or consumption of the system participants. More details about the SIMONA simulation environment can be found in [106], [107]. Utilizing the developed model of the electrical grid infrastructure and the system participants (see Section 3.2), a one-year simulation run is performed. This data set is henceforth referred to as the synthetic data set, modeling the behavior of the MV grid and leveraged for supervised training of the ML models

3.3.2 Validation

To assess the validity of the simulated data set, the simulation are compared with the corresponding measurements. No further fine tuning of load or generation parameters has been performed for two reasons. Firstly, most available data of the DG operator about load and generation systems will be imprecise. Any individ-

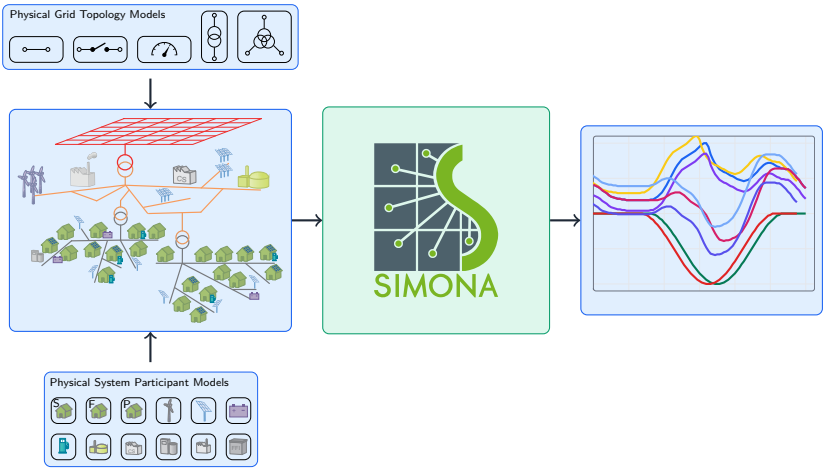


Figure 3.11: Overview of power system simulation based on [106].

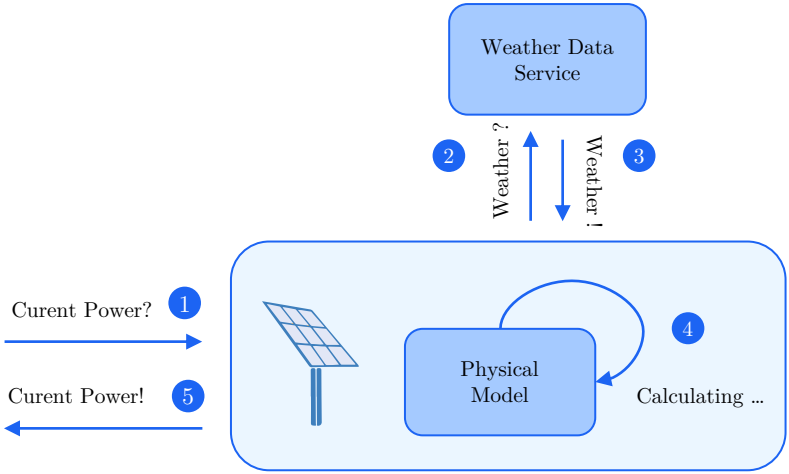


Figure 3.12: Exemplary PV model calculation within the SIMONA simulation environment.

ual tinkering with values can not be automated and is hard to reproduce. For the supervised approach to be feasible, the model must be trainable with an imperfect representation of the real data distribution. Secondly, manual adjustments based on deviations during validation would only be possible for the measured buses. Since the performance evaluation uses these measured buses as proxies for unmeasured buses, any fine-tuning specific to measured buses could artificially improve results at those locations. This improvement would not be reproducible for unmeasured buses, potentially leading to an overestimation of the model's overall performance.

A comparison of the respective voltage distributions is shown in Figure 3.13a. It can be seen that the overall distributions are similar. The buses with higher numbers are slightly skewed in the negative direction. As bus numbering starts at the slack bus, these are buses deeper in the grid this indicates slightly higher load in the synthetic data. Figure 3.13b depicts a comparison between bus active power distributions. Given the limited amount of information available for configuring the load and generation models, the active power distribution is reasonably similar in shape.

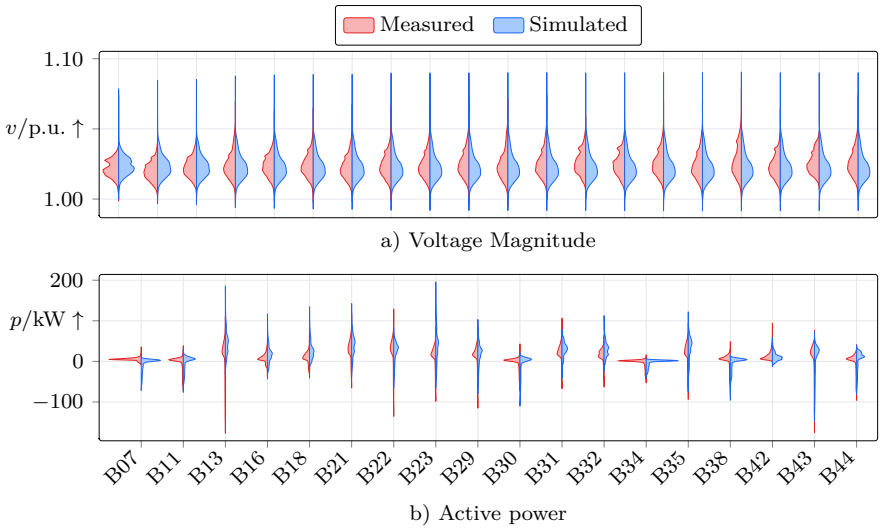


Figure 3.13: Comparison of simulated and measured bus voltages.

3.4 Summary

This chapter presents the development and validation of a synthetic grid model intended for generating training data for data-driven DSSE approaches. Initially the grid that forms the basis for the DSSE is described, and its graph structure is presented. The grid primarily serves residential loads with a significant share of distributed generation, predominantly from photovoltaic systems.

Examining of field measurements reveals several challenges regarding data quality that need to be considered. These include systematic measurement errors like signal losses and unregistered tap changes, which require the development of appropriate filtering methods. Analysis of measurement noise suggests that some form of measurement preprocessing was applied to the raw measurements, although specific details are unavailable.

Based on the physical grid, a synthetic model is developed by establishing a conversion workflow to transform the real grid model into a simulation-ready format. The modeling of system participants focuses on the three dominant types present in the grid: households, photovoltaic systems, and night storage heating. Individual participants are modeled based on a data source provided by the grid operator that denotes basic bus-specific power values and quantities for the different system types. Households are modeled stochastically with an assumed yearly consumption drawn from a standard distribution with a mean of 3.500 kWh of energy consumption per year. PV systems are modeled based on known bus-specific power and quantity. Additional parameters like elevation angle and orientation are sampled according to an empiric distribution. Night storage heating is implemented using temperature-dependent load profiles scaled to the rated power denoted for each bus.

The synthetic model forms the basis for a time series simulation of one year using the energy system simulator SIMONA. The validation of the simulation results shows that it produces voltage distributions similar to the measured data, with a slight negative skew for buses further from the slack bus. Bus-specific active power distributions show reasonable alignment, considering the limited information about load and generation systems. No special fine-tuning of load and generation systems is performed to maintain generalizability and to reflect the realistic scenario where precise system information is limited. This approach ensures that the subsequent training of DSSE models reflects real-world conditions where a perfect representation of the underlying grid state cannot be assumed.

4 Handling of Missing and Anomalous Measurements

The synthetic NSE approaches neither answer the question of how to employ the model in the field, nor do they have to contend with the inconveniences of real measurement data. Measurements by nature are imprecise to different degrees, which is acknowledged by the existence of measurement accuracy classes as discussed in Section 2.2. Additionally, there is no such thing as perfect reliability of the measurement device or the signal transmission, leading to signal loss or faulty transmissions. Furthermore, a mismatch between information about the system state, like unregistered tap changes, can lead to voltage jumps that are not accounted for. Both scenarios were found in the measurement data set used in this thesis and shown in Figure 3.2. Those types of measurement errors can be categorized as measurement anomalies. Any anomalous measurement in the context of grid operation means wrong values are being displayed for the state of critical systems. If undetected, those anomalous measurements can lead to a cascade of errors in downstream applications.

These issues can be alleviated through anomaly detection techniques and subsequent signal reconstruction, which are the focus of this chapter. For both tasks the Autoencoder ANN architecture is leveraged. The model is trained on the synthetic data set, and its transferability to the real measurement data is analyzed. Before diving into the specific mechanisms of anomaly detection and signal reconstruction, the general model training and ML-specific data preprocessing steps are described.

4.1 Model Training

The basic procedures of the model training process are outlined in Figure 4.1. The *data set* is comprised of the entirety of the data that could be utilized in the training process. In the application case, this is represented by the simulated grid state. *Feature selection* in the context of this thesis is defined by the selection of measurement locations in the grid structure and the selection of electrical attributes that are measured at the respective locations. During the *training data split*, the dataset is divided into training, validation, and test data sets. The training data set is utilized for actual model training. The validation data set is employed for model performance evaluation during hyperparameter optimization, and the test set is reserved for final model evaluation. The specifics of the data-splitting process are described in Section 4.2. Since neural networks have been proven to learn most effectively with input vectors containing values between zero and one [60], *feature scaling* is

implemented. Min-max scaling is selected as the scaling procedure. Given the variety of hyperparameters, such as number of layers, number of neurons, activation function and optimizer among others, *hyperparameter tuning* is conducted to determine the optimal parameter configuration. The ASHA scheduler of *RayTune* [108] is used for effective hyperparameter search, which is based on [109]. Early stopping is implemented throughout model training to prevent overfitting.

Figure 4.1: General model training pipeline.

4.2 Stratified Date Sampling

Power distribution networks exhibit strong temporal patterns that must be considered when preparing data for machine learning applications. Both electricity consumption and renewable generation show significant variations across different seasons and days of the week. For instance, PV generation is heavily influenced by weather conditions throughout the year, while electricity consumption patterns vary based on both weather conditions and weekday types. During colder months, consumers typically use more electricity for heating and spend more time indoors, leading to distinct consumption patterns. These types of seasonal and weekly differences are also represented in the standard load profiles [110].

To develop robust machine learning models, it is crucial to ensure that the training, validation, and test datasets fairly represent these diverse temporal patterns. A naive sequential split of the data shown in Figure 4.2a would not capture the full range of seasonal variations in each dataset. Therefore, a stratified sampling algorithm that accounts for seasonal and weekly patterns is implemented. A visualization of the resulting split is depicted in Figure 4.2 b). The remainder of this Section outlines the developed sampling algorithm.

Let D be the complete dataset of measurements. The sampling strategy categorizes data points based on two key attributes: seasons $s \in$

$S = \{\text{Spring}, \text{Summer}, \text{Autumn}, \text{Winter}\}$ and day types $k \in K = \{\text{Weekday}, \text{Saturday}, \text{Sunday}\}$. This categorization results in twelve distinct combinations that capture the various temporal contexts present in the data. For each combination (s, k), its relative frequency in the complete dataset is defined as:

$$\eta_{s,k} = \frac{|\{x \in D : \text{season}(x) = s \wedge \text{daytype}(x) = k\}|}{|D|} \tag{4.1}$$

where $|\cdot|$ denotes the cardinality of a set. The sampling creates three disjoint sets: training set D_{training}, validation set D_{valid}, and test set D_{test}, such that $D_{\text{training}} \cup D_{\text{valid}} \cup D_{\text{test}} = D$ and $D_{\text{training}} \cap D_{\text{valid}} = D_{\text{valid}} \cap D_{\text{test}} = D_{\text{training}} \cap D_{\text{test}} = \emptyset$.

The key feature of this sampling approach is that it maintains the relative distribution of the season-day type combinations across all three datasets. Mathematically, the following is true for each combination (s, k):

$$\begin{aligned} \frac{|\{x \in D_{\text{training}} : \text{season}(x) = s \wedge \text{daytype}(x) = k\}|}{|D_{\text{training}}|} &\approx \eta_{s,k} \\ \frac{|\{x \in D_{\text{valid}} : \text{season}(x) = s \wedge \text{daytype}(x) = k\}|}{|D_{\text{valid}}|} &\approx \eta_{s,k} \\ \frac{|\{x \in D_{\text{test}} : \text{season}(x) = s \wedge \text{daytype}(x) = k\}|}{|D_{\text{test}}|} &\approx \eta_{s,k} \end{aligned} \tag{4.2}$$

For example, if saturdays in spring represent 10% of the total data, they will also constitute approximately 10% of each data subset (D_{training}, D_{valid}, and D_{test}).

A downside of this approach is that the resulting samples of the different sets can be non-consecutive. Since some ML models rely on input sequences like long short-term memory ANNs this would be problematic. Therefore, instead of considering each system state snapshot as a sample for the sampling approach, the considered samples comprise all consecutive snapshots of the day. This allows model training with sequence length of up to a whole day, while preserving the opportunity for stratified sampling.

4.3 Anomaly Detection

In this chapter a model is presented to detect anomalous signals. The detection mechanism can also be exploited to reconstruct the signal and improve the overall system state observation. The signal reconstruction will be discussed in Section 4.4.

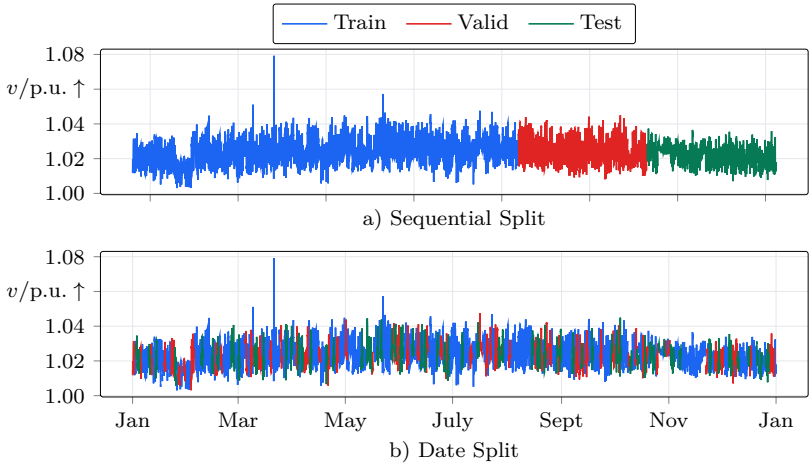


Figure 4.2: Exemplary training (60 %), test (20 %), validation (20 %) data split for the voltage magnitude.

4.3.1 Utilizing Autoencoders for Anomaly Detection

To perform anomaly detection, we leverage the autoencoder architecture. The underlying assumption is that the autoencoder is able to reconstruct valid data well but will produce larger reconstruction errors for anomalous data. Given that assumption, the reconstruction error Δ_{x_i} of feature x_i can indicate the feature's validity. The reconstruction error is the delta between feature x_i and its reconstruction $\tilde{x}_i$:

$$\Delta_{x_i} = x_i - \tilde{x}_i \tag{4.3}$$

If Δ_{x_i} exceeds a set threshold τ the feature x_i is considered an anomaly. The process is visualized in Figure 4.3. The threshold τ that determines whether a sample is considered an anomaly is selected by choosing the highest F1 score across different threshold values. The F1 score is the harmonic mean of precision and recall. Precision measures the proportion of correctly identified anomalies among all predicted anomalies. Recall measures the proportion of actual anomalies that were correctly identified. Each individual reconstruction error Δ_{x_i} serves as a potential threshold candidate τ. For each candidate threshold, the F1 score is calculated using the precision and recall values obtained by classifying samples as anomalous when

their reconstruction error exceeds the threshold. The optimal threshold τ^* is then selected as the one that maximizes the F1 score:

$$\tau^* = \text{argmax } \text{F1}(\tau) \tag{4.4}$$

This approach provides a balance between over-sensitive detection with excessive false positives and under-sensitive detection that misses too many anomalies.

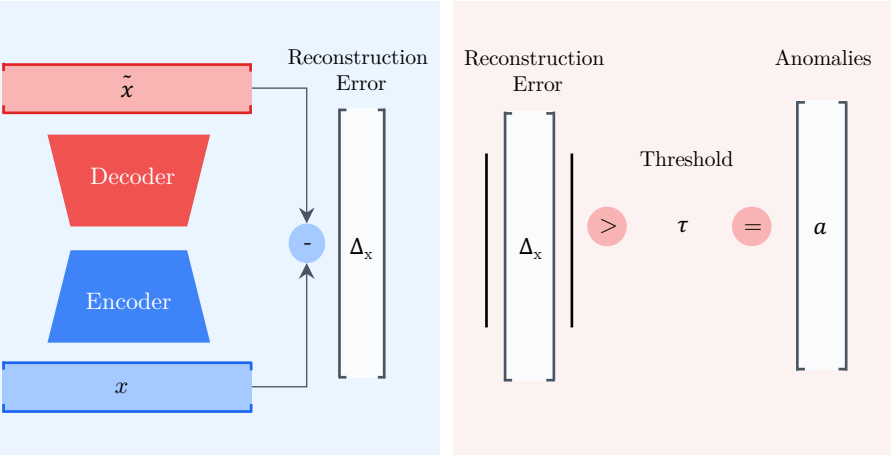


Figure 4.3: Anomaly detection mechanism consisting of calculation of reconstruction error (left) and threshold based detection of anomalies (right).

4.3.2 Training and Evaluation

The encoder and decoder architectures were implemented as symmetrical Multilayer Perceptrons (MLPs). The general training procedure used was outlined in Section 4.1. The input and output vector of the model is made up of the voltage magnitudes vector $\vec{v}$ of all measured buses. For the hyperparameter tuning, 200 runs were performed, resulting in the final hyperparameter configuration outlined in Table 4.1.

Table 4.1: Resulting autoencoder hyperparameters after hyperparameter tuning.

Hyperparameter	**Setting**
Number of MLP layers	4
Layer size	150
Activation Function	ReLU
Loss Function	MSELoss
Learning Rate	1.54×10^{-3}
Optimizer	Adam
Batch Size	32

To test and evaluate the model, anomalies are introduced into the synthetic data. Anomalies are introduced by randomly adding a voltage deviation to the true voltage $\vec{v}$ as outlined in Algorithm 1.

Algorithm 1: Introduce Synthetic Anomalies

Input: Voltage matrix $\mathbf{V} \in \mathbb{R}^{|B|\times|T|}$, error rate $\varepsilon \in \mathbb{R}^{+}$
Output: Voltage matrix with anomalies $\mathbf{V}^{\text{anom}} \in \mathbb{R}^{|B|\times|T|}$

```
for i ∈ B do
    for t ∈ T do
        if rand() < 0.05 then
            δ ← v_{i,t} × ε;
            if rand() < 0.5 then
                v^anom_{i,t} ← v_{i,t} + δ;
            else
                v^anom_{i,t} ← v_{i,t} − δ;
        else
            v^anom_{i,t} ← v_{i,t};
```

An exemplary time series with an error rate of 0.05 and the corresponding anomaly detection using a day of the synthetic test data is illustrated in Figure 4.4. Subplot a) visualizes the original time series and one with synthetically added anomalies. Each red dot constitutes a synthetic anomaly. In Subplot b), the reconstruction error Δ_x for each of the values of the anomalous signal is visualized. Given the

calculated optimal threshold of about 0.07, visualized by the red striped line, all anomalies can be separated from the valid values and are thereby correctly identified as anomalies.

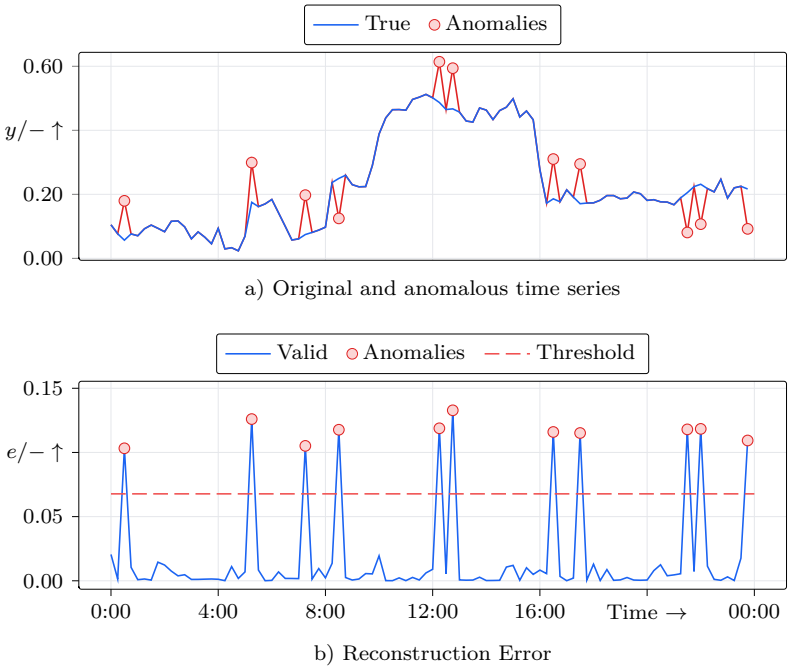


Figure 4.4: Anomaly Detection on synthetical errors.

To assess the ability of the model to detect anomalies with different error magnitudes, tests were done using error rates between 0.3 % and 5 % and a step size of 0.1 %. Using Algorithm 1, for each error rate anomalies were introduced to the test set. For each of the 47 resulting data sets, the respective recall and precision is depicted in Figure 4.5. Even for the lowest error rates the recall exceeds 0.94 and the precision 0.988 illustrating the model's effectiveness for the synthetic data set.

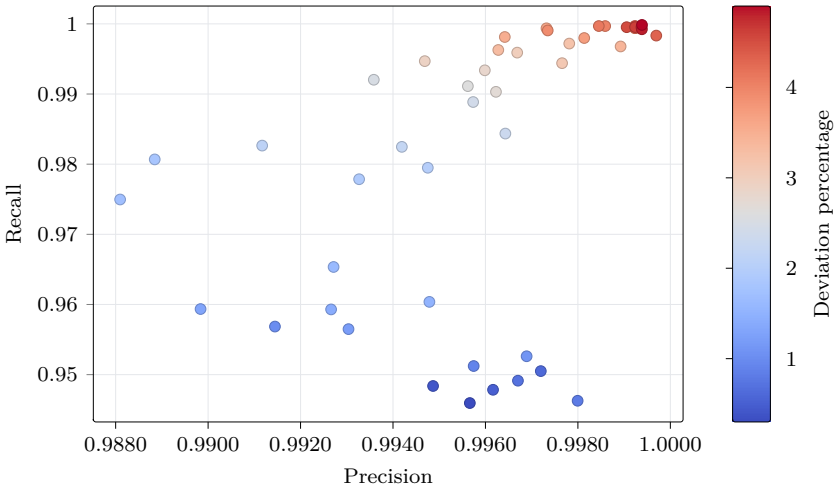


Figure 4.5: Precision versus recall scatter plot for different error rates.

4.3.3 Anomaly Detection on Measurement Data

Although the model showed promising results on synthetic data, transferability to real-world application must be demonstrated. Therefore, the model's performance on actual measurement data is investigated in the following section.

Finding the best threshold previously was based on having labels for the data set. However, the measurement data is unlabeled, with no information available to determine whether or not any measurement is corrupt. Consequently, another method for calculating the threshold is necessary.

The reconstruction error on the measurement data set was calculated and is visualized in a box plot in Figure 4.6. The IQR of the reconstruction error is about 0.0015. Given that the outliers get as large as 0.1562, which is about 100 times the IQR, this indicates a clear distinction between those specific measurements and the rest of the data set. This distinction can be assumed to be due to measurement errors which hints at the general applicability of the model to the measurement distribution. To balance sensitivity to anomalies and robustness against false positives, the threshold was chosen to be the 99th percentile reconstruction error value. Thus, the 1 % of measurements exceeding that threshhold are considered anomalous.

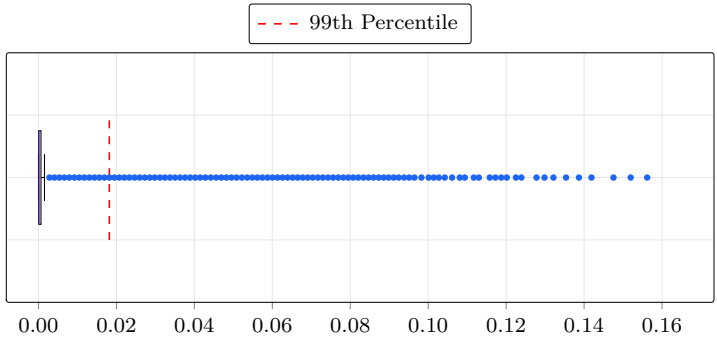


Figure 4.6: Reconstruction errors of Deep Linear Autoencoder (DLAE) model output, given the measurement data.

An exemplary application of anomaly detection using measurement data is depicted in Figure 4.7. Since the ground truth of potentially anomalous measurements is unknown, the analysis must be performed qualitatively. Similarly, as in Figure 3.2, the signal was plotted along with the closest surrounding measurements. The investigated measurement shows a jump in voltage difference to the surrounding measurements 0:00 and 03:00 and 12:30 to 17:30. This suggests a tap change of the MV/LV transformer, which was not accounted for. As visualized in the subplot b), the noticeable segments show a reconstruction error above the threshold and are, therfore, labeled anomalous. The chosen threshold was effective for this data set. In general more research needs to be done that evaluates different mechanisms by which to choose threshold values for unlabeled data sets.

4.4 Measurement Reconstruction

Given the potential grave implications of anomalous measurements, and the demonstrated ability to find those anomalies as discussed in Section 4.3, a way to deal with anomalies is necessary. One option is to drop the anomalous data. This would leave us with worse observability of the system, which is already subpar in today's DGs. A second option is the reconstruction of the anomalous signal, given the healthy information about the system.

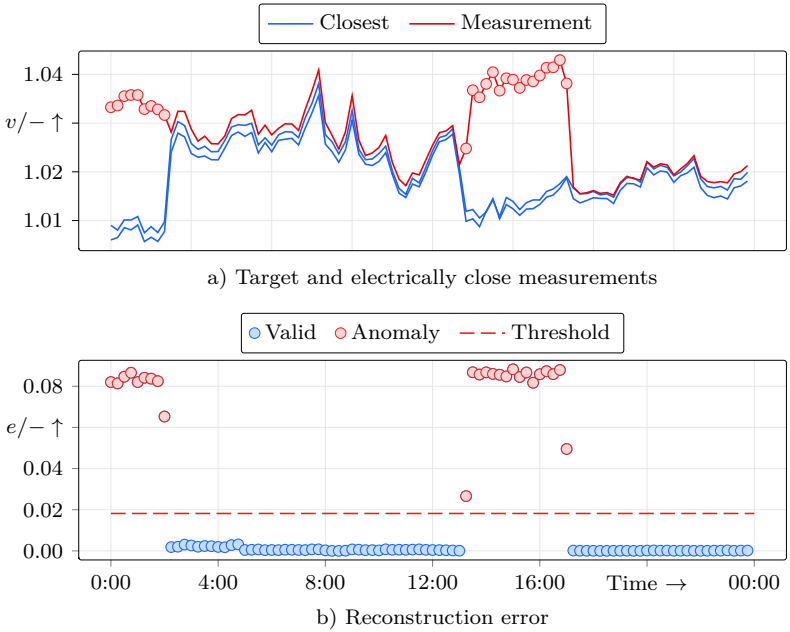


Figure 4.7: Anomaly detection for exemplary day, including electrically close measurements for reference.

For measurement reconstruction, the same DLAE model also employed for anomaly detection can be utilized. The procedure is made up of two steps. First the anomalous signals are substituted by pseudo-measurements which are generated on the basis of the remaining healthy signals Secondly, the pseudo-measurements are shifted by utilizing the DLAE reconstruction. Both steps will be described in this chapter in more detail.

4.4.1 Calculating Pseudo-Measurements

Due to the correlation of adjacent nodal voltages, the closest measurements of the anomalous signal will generally be most similar to the anomalous signal. The proposed algorithm traverses the graph and finds the closest valid measurements. The

pseudo-measurement is then made up of a weighted sum of those valid measurements.

Given the anomalous signals $b_{\text{i, anomalous}} \in B_{\text{anomalous}}$, the set of valid buses with corresponding measurements is B_{valid}.

$$B_{\text{valid}} = B_{\text{measured}} \setminus B_{\text{anomalous}} \tag{4.5}$$

For each missing signal, the graph is traversed in each direction. When visiting a bus b_i, and the bus is not part of B_{valid}, the traversal continues, visiting all of the neighbors of b_i that have not yet been visited. If $b_i \in B_{\text{valid}}$ or b_i has no further neighbors, the traversal stops. All buses found during graph traversal comprise B_{pseudo} and are considered for calculating the pseudo-measurement. An exemplary graph traversal is shown in Figure 4.8. The buses $B_{\text{pseudo}} = \{b_1, b_4, b_5\}$ are the basis for pseudo-measurement generation in the given example.

Afterwards, the measurements of B_{pseudo} are weighted inversely with respect to their distance from the anomalous bus. All buses in B_{pseudo} are categorized into upstream buses B_{upstream} and downstream buses $B_{\text{downstream}}$. The Djikstra shortest path distance [111] between b_i and b_j is denoted as $d(b_i, b_j)$. A bus is considered upstream if its shortest path distance to the slack bus is shorter than the distance of the anomalous bus to the slack bus.

$$\begin{aligned} B_{\text{upstream}} &= \{b_i \in B_{\text{valid}} : d(b_i, b_{\text{slack}}) < d(b_{\text{anomalous}}, b_{\text{slack}})\} \\ B_{\text{downstream}} &= \{b_i \in B_{\text{valid}} : d(b_i, b_{\text{slack}}) \geq d(b_{\text{anomalous}}, b_{\text{slack}})\} \end{aligned} \tag{4.6}$$

For the example shown in Figure 4.8, this categorization yields $B_{\text{upstream}} = \{b_1\}$ and $B_{\text{downstream}} = \{b_4, b_5\}$. Both groups are weighted by the inverse of the shortest distance of their respective buses to the anomalous bus. The distances are noted as $d_{\text{min}}^{\text{downstream}}$ and $d_{\text{min}}^{\text{upstream}}$.

The group's weighting factors are then described as:

$$\begin{aligned} w_{\text{downstream}} &= \frac{d_{\text{min}}^{\text{upstream}}}{d_{\text{min}}^{\text{downstream}} + d_{\text{min}}^{\text{upstream}}} \\ w_{\text{upstream}} &= \frac{d_{\text{min}}^{\text{downstream}}}{d_{\text{min}}^{\text{downstream}} + d_{\text{min}}^{\text{upstream}}} \end{aligned} \tag{4.7}$$

Within the groups B_{upstream} and $B_{\text{downstream}}$, each measurement is additionally inversely weighted by its minimal distance to the anomalous bus. If we define

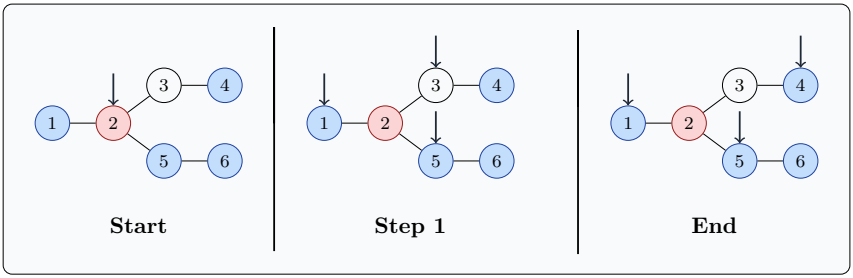


Figure 4.8: Graph traversal illustration.

the inverse distance of bus b_i to the bus with the anomalous signal $b_{\text{anomalous}}$ as $d^{-1}_{\text{i, anomalous}}$, the weights can be calculated with

$$w_i = \frac{d^{-1}_{\text{i, anomalous}}}{\sum_{j:b_j \in B_k} d^{-1}_{\text{j, anomalous}}}, \quad \forall i : b_i \in B_k \tag{4.8}$$

$$\sum_{i:b_i \in B_k} w_i = 1 \tag{4.9}$$

where $B_k \in \{B_{\text{upstream}}, B_{\text{downstream}}\}$. The final pseudo-measurement $x_{\text{i, pseudo}}$ of bus $b_i \in B_{\text{pseudo}}$ can then be calculated as follows:

$$\begin{aligned} x_{\text{i, pseudo}} = & \sum_{j:b_j \in B_{\text{upstream}}} w_{\text{upstream}} \times w_j \times x_j + \\ & \sum_{l:b_l \in B_{\text{downstream}}} w_{\text{downstream}} \times w_l \times x_l \end{aligned} \tag{4.10}$$

Replacing the anomalous signals with calculated pesudo-measurements, as outlined above, can work sufficiently well when measurements surround an anomalous signal. However, suppose there are only measurements in either the downstream or upstream direction, and their distance to the anomalous signal is large. In that case, the pseudo-measurement will have a large offset in either positive or negative direction. This effect is particularly evident when the slack voltage signal is missing, as shown in Figure 4.9 for an exemplary day using the synthetic dataset. Since there is, by definition, no measurement upstream of the slack voltage signal, and the closest downstream measurements are not close to the slack bus (see Figure 3.1), a

significant offset between the actual measurement and the pseudo-measurement is expected.

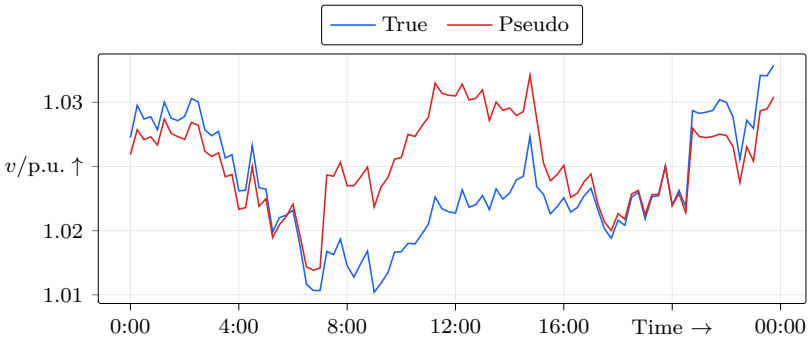


Figure 4.9: Pseudo-measurement for a missing slack voltage signal.

4.4.2 Utilizing Autoencoders

The performance of pseudo-measurement generation for signal reconstruction can be enhanced by leveraging the DLAE model. The DLAE's ability to correct imprecise measurements stems from its training on reconstruction tasks, where it learns underlying patterns in power system measurements.

The autoencoder-based reconstruction process is initiated by replacing the anomalous signals with pseudo measurements generated using the method described in the previous section. This initial replacement can be expressed as:

$$x_i = \begin{cases} x_i, & \text{if } i \notin A \\ x_{i,\text{pseudo}}, & \text{if } i \in A \end{cases} \tag{4.11}$$

where A is defined as the set of indices corresponding to anomalous signals, and $x_{i,\text{pseudo}}$ is the pseudo-measurement for the i-th anomalous signal. The DLAE is then applied to the adjusted input vector $\vec{x}$. The DLAE can be represented as a function $\mathbf{D}$, such that:

$$\vec{y} = D(\vec{x}) \tag{4.12}$$

where $\mathbf{y}$ is defined as the output vector of the DLAE. The output y_i for positions ($i \in A$) is considered the new measurement reconstruction. The pseudo-measurements are effectively shifted towards more probable values based on the DLAE's learned representation of the system. The corresponding y of the model given the anomalous slack bus example from Figure 4.9 is visualized in Figure 4.10.

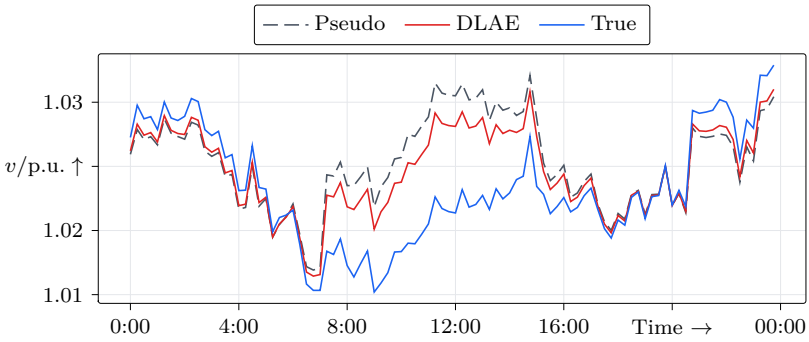


Figure 4.10: Measurement reconstruction after DLAE application.

While the model does reduce the error, a single pass through the model does not sufficiently shift the pseudo-measurements. Therefore, a procedure is implemented to iteratively update the input vector x. In each iteration the DLAE gets reapplied to the updated input by replacing the values $x_i : i \in A$ with the corresponding outputs from the DLAE. Thus, the input in iteration r is:

$$x_i^r = \begin{cases} x_i, & \text{if } i \notin A \\ y_i^{r-1}, & \text{if } i \in A \end{cases} \tag{4.13}$$

with y_i^{r-1} being the model output of the iteration $r-1$. To ensure convergence and avoid unnecessary iterations, two stopping criteria are implemented:

Convergence criterion: The difference between the model's output and the model's input at the anomalous signal positions is calculated:

$$\delta = \max_{i \in A}(|y_i - x_i|) \tag{4.14}$$

The iterative procedure is stopped when:

$$\delta \leq 10^{-6} \tag{4.15}$$

Maximum iteration criterion: Given the iteration count c, an upper limit on the number of iterations is set to prevent excessive computational time:

$$c \leq 100 \tag{4.16}$$

The updated measurement reconstruction using the previously described iterative approach, applied to the exemplary scenario, is depicted in Figure 4.11. The summarized reconstruction procedure is depicted in Figure 4.12.

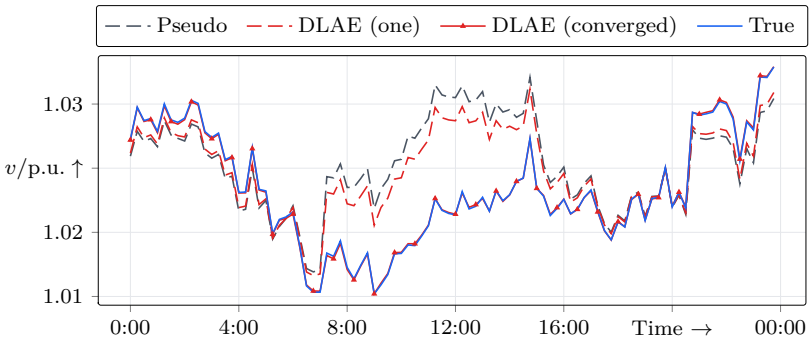


Figure 4.11: Measurement reconstruction with iterative DLAE application on synthetic dataset.

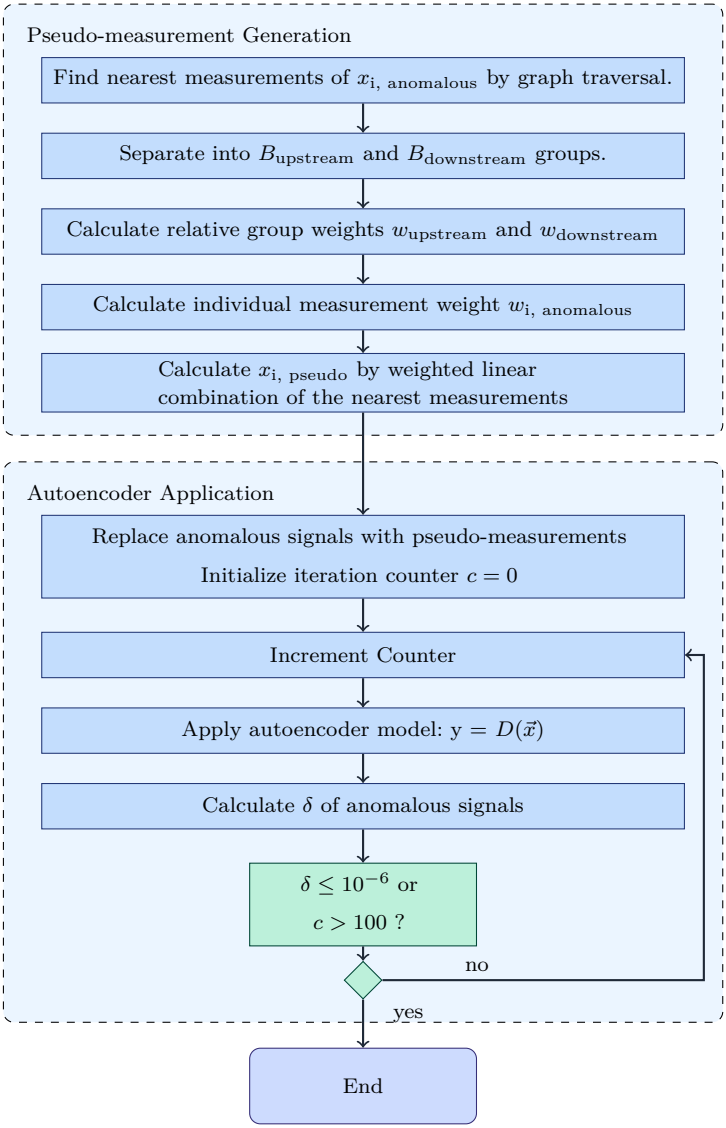


Figure 4.12: Anomalous measurement reconstruction procedure.

4.4.3 Performance Evaluation

To assess the real-world performance of the measurement reconstruction and the generalizability of the DLAE, its accuracy on the measurement data is investigated.

In order to evaluate the reconstruction accuracy, the measurement reconstruction is run for a range of different numbers of missing measurements. Given the geographical distribution of measurements in the grid, not only the number of missing measurements is expected to play a significant role in the reconstruction accuracy, but also their location. To get a sense of the average performance of a certain number of missing measurements, the location of the missing measurements is resampled every day. A flow chart on how the reconstruction is evaluated on the measurement data is depicted in Figure 4.13.

Figure 4.14 visualizes the resulting reconstruction deviation. As expected, the voltage deviation increases as more signals have to be reconstructed, and simultaneously, the number of valid signals off of which the signals can be reconstructed decreases. As visible in Figure 4.14a, the average performance is still relatively accurate for higher ranges of missing signals. This is partly explained by the relatively low deviation between the different nodal voltages when load and generation in the system are comparatively low. Since there is little current flowing in these situations, the voltage drop across the lines is small. Thus, the measured signals and derived pseudo measurements are similar in value, leading to little corrections by the DLAE reconstruction and little deviations overall. For an increasing amount of missing measurements, outliers become more frequent and more extreme, and there is a relatively strong performance degradation when only two measurements are left. While outliers for one missing signal are below a 1 % deviation, they exceed 2 % and even 3 % deviation for higher numbers of missing signals. For practical purpose a minimal performance characteristic is likely required. Accordingly a threshold of 0.5 % is set based on the more stringent DSSE success criterion defined in [65]. To indicate the impact of the performance degradation with respect to the set threshold, Figure 4.14b depicts the relative amount of reconstruction deviation exceeding 0.5 %. The amount of reconstructions exceeding the threshold is dented as the failure rate FR. A roughly linear increase from 5 % to 9 % with a sudden jump from 9 % to 16 % at 17 missing signals. Overall, the method proves effective in reconstructing a variable amount of missing measurements. Specific evalutation of the reconstruction performance has to be done with respect to the requirements of the downstream application of the reconstructed measurements.

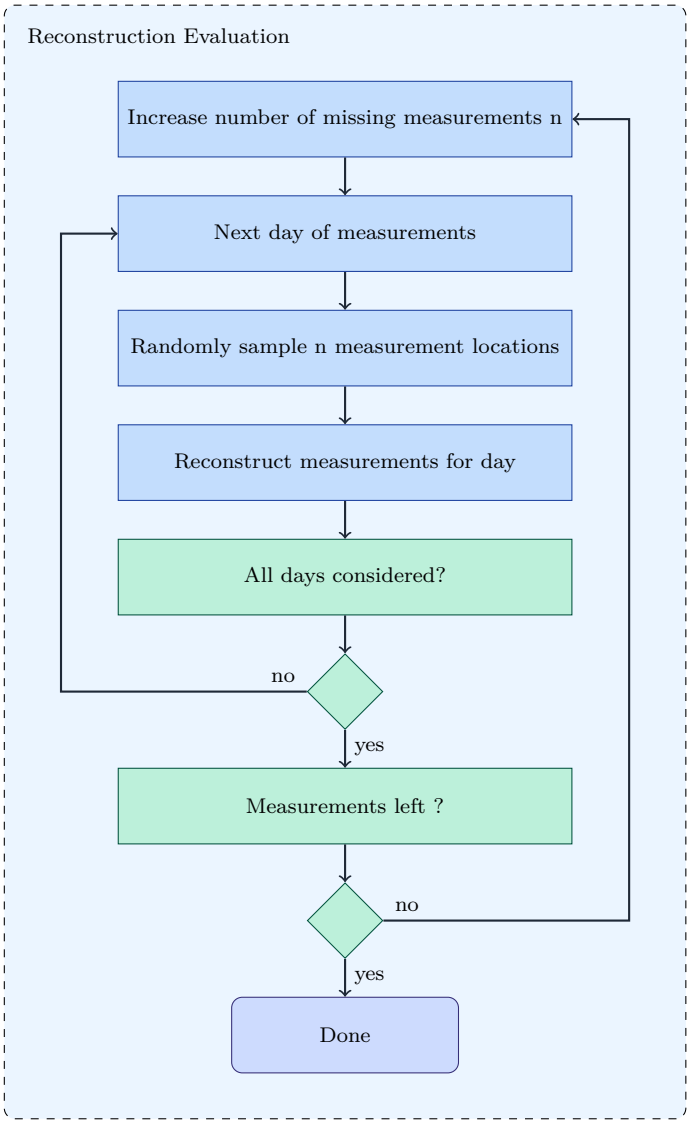


Figure 4.13: Evaluation process for reconstruction on the measurement data set.

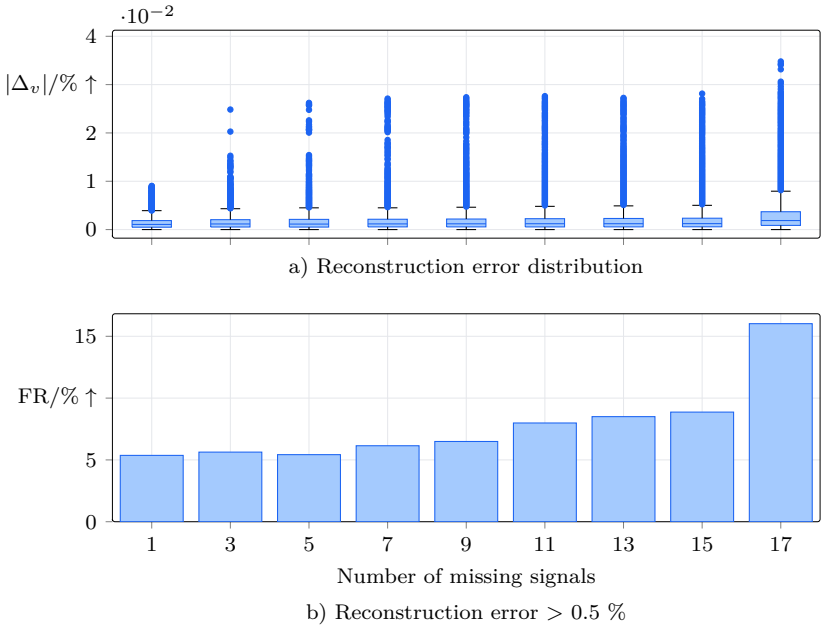


Figure 4.14: Distribution of voltage deviation between measurement and reconstructed signal across different numbers of missing signals.

4.5 Summary

Chapter 4 investigates approaches for handling missing and anomalous measurements in power distribution grid monitoring, leveraging supervised learning. Initially the steps and procedure of model training is described. The general training pipeline incorporates feature selection, data splitting, min-max scaling, and hyperparameter optimization. A stratified sampling approach is presented for data splitting, that accounts for the temporal characteristics of power systems, considering both seasonal variations and weekly patterns. This sampling strategy categorizes data points based on seasons and day types, maintaining their relative frequencies across training, validation, and test sets. The implementation allows for consecutive

daily samples, facilitating the training of sequence-based models while preserving the benefits of stratified sampling.

The first step in handling anomalous data is their detection. For anomaly detection, an autoencoder-based architecture is investigated. The approach leverages the autoencoder's ability to reconstruct valid data with small errors while producing larger reconstruction errors for anomalous data. Through hyperparameter tuning involving 200 runs, an architecture with four MLP layers and 150 neurons per layer is established. The model's performance is evaluated on synthetic data with varying error magnitudes ranging from 0.3 % to 5 %. The subsequent analysis shows recall rates exceeding 0.94 and precision above 0.99 across all error ranges, indicating the model's capability to identify anomalies effectively. The transferability to real measurement data is examined, requiring adaptation of the threshold selection process due to the absence of ground truth labels. Analysis of the reconstruction error distribution on measurement data shows a narrow interquartile range. This indicates an overall low reconstruction error with some outliers, indicating potential measurement errors. The 99th percentile is selected as the detection threshold to balance sensitivity and robustness. Previously suspected measurement anomalies were successfully identified by the model, suggesting its applicability to the measurement data.

Anomalous or missing data in the context of grid monitoring can be problematic in downstream applications. Based on the anomaly detection model a signal reconstruction method is analyzed. The method is made up of two separate steps. Firstly, pseudo-measurements are generated based on graph traversal and weighted averaging of valid neighboring measurements. The method categorizes measurements into upstream and downstream groups, applying distance-based weighting to calculate replacement values. Afterwards, these initial estimates are refined through iterative application of the autoencoder model used for anomaly detection until convergence criteria are met. Performance evaluation of the reconstruction method is conducted using measurement data with varying numbers of missing signals. The analysis demonstrates the method's capability to handle multiple simultaneous reconstructions, though accuracy decreased with an increasing number of missing measurements. On average the reconstruction performance was still good for higher numbers of missing measurements, partially attributable to low deviations between underlying bus voltages when current flows are low. Notably, outliers get more extreme for higher numbers of missing measurements reaching deviations exceeding 3 %.

5 Neural State Estimation

Applying supervised learning techniques to the DSSE problem neccessitates a complete data set not available in sparsely measured systems. While literature approaches relied on synthetic data sets, no research was done on how transferable these models are to field applications. In this chapter, an NSE model is trained on the synthetic data set created in Chapter 3. It is evaluated against a WLS implementation enhanced by pseudo-measurements on synthetic data. Afterwards, its expected performance in a real-world scenario is quantified by evaluating it with respect to the field measurements of the grid. The model's robustness to changes in load and generation is evaluated and fine-tuning is proposed to improve its robustness. Furthermore, Monte Carlo (MC) Dropout is proposed for quantifying the model's estimation uncertainty.

5.1 WLS Reference Model

To better evaluate the performance of the NSE a benchmark model is necessary. For that matter, a WLS-based SE is implemented. The process of retrieving the measurements for the WLS DSSE is summarized in Figure 5.1. Given the retrieved measurements, the process of executing the DSSE is depicted in Figure 5.2.

As the computational basis, the SE core of pandapower[3] is used. Hence, the grid model is converted to a pandapower grid. Given the nature of the WLS algorithm, pseudo-measurements are needed to make the grid model observable. For every unobservable node, active, and reactive power values are generated. As the basis for the generation the system participant models, described in Section 3.2.2, are utilized. For the household load, the stochastic load model is replaced with standard load profiles, scaled to the participants' yearly energy consumption. For realistic modeling of PV generation and night storage heating demand, weather data for the specific time interval of the measurements is necessary. To that end, the ERA5 reanalysis data set for 2023 is utilized [112]. The input of the WLS state estimation comprises the voltage magnitude and power values for each measured node, as well as the active and reactive power of each unmeasured node. As for parametrization, the maximum iteration count of the algorithm is set to ten, which is the default value. Given that the algorithm might not converge for tight convergence criteria, given the limited amount of data points and the potentially high error of pseudo-

[3] pandapower on Github - https://github.com/e2nIEE/pandapower

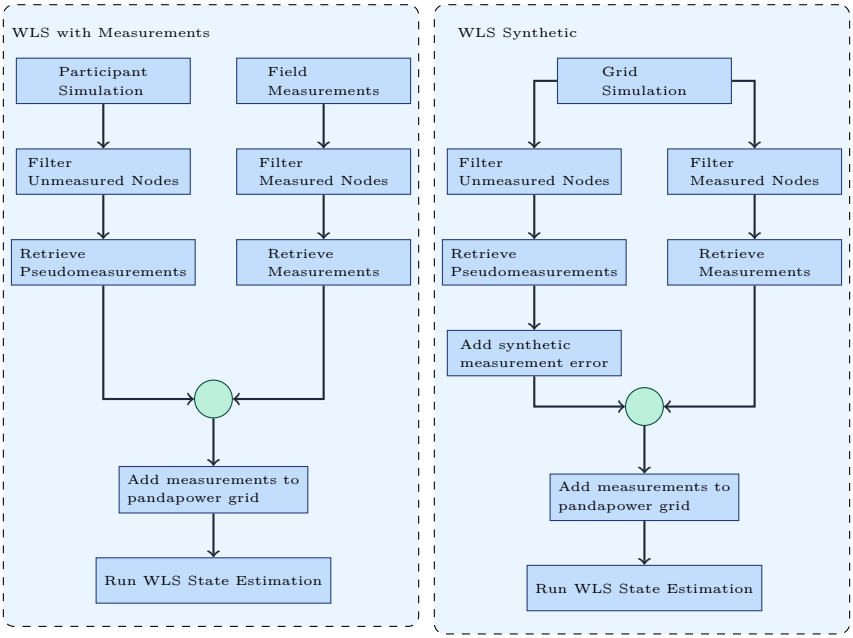


Figure 5.1: Extraction and generation of measurements and pseudo-measurements for the WLS DSSE. The procedures are separated into the measurement and synthetic scenario.

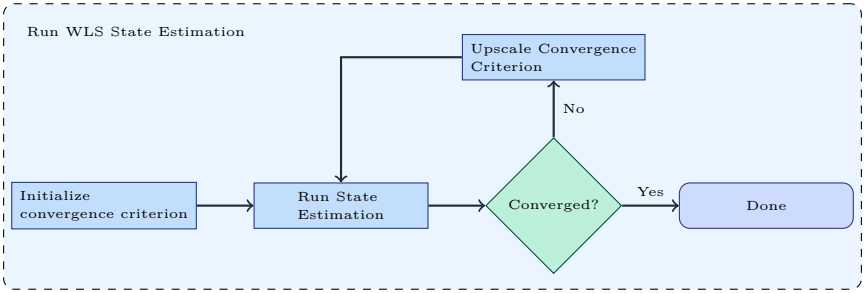


Figure 5.2: WLS DSSE procedure with proportional backoff.

measurements, a proportional backoff is implemented for the criterion's magnitude. Starting with $1e^{-6}$, its magnitude gets doubled upon each unsuccesful iteration until the algorithm converges.

While this methodology works for the generation of pseudo-measurements when utilizing the measurement as inputs, it does not for the synthetic data set. As the model calculations are the underlying ground truth values for the synthetic data set, using them as pseudo-measurement would underestimate typical error distributions of pseudo-measurements. To simulate the estimation error of the pseudo-measurements, noise is added to the actual active and reactive power values. An exemplary active power time series with synthetically generated noise is depicted in Figure 5.3.

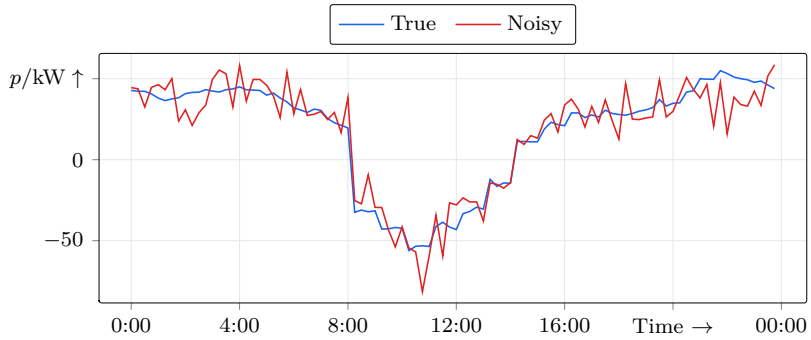


Figure 5.3: Power time series and corresponding adjusted noisy time series for usage as pseudo-measurement.

The noise is sampled from a Gaussian distribution that generates Relative Average Deviation (RAD) of 20 %. Given the RAD, the Absolute Average Deviation (AAD) of a value x_i is calculated as:

$$\text{AAD} = \text{RAD} * x_i \tag{5.1}$$

For normal distribution, the standard deviation and absolute average deviation are correlated by a factor of rougly 0.7979. Hence, the standard deviation for the Gaussian distribution is calculated as follows:

$$\sigma_{\text{pseudo, i}} = \frac{\text{AAD}}{0.7979} \tag{5.2}$$

The resulting "pseudo-measurement" $x_{\text{pseudo,i}}$ is:

$$x_{\text{pseudo,i}} = x_i + N(0, \sigma_{\text{pseudo, i}}) \tag{5.3}$$

where $N(0, \sigma_{\text{pseudo, i}})$ is the Gaussian error distribution with zero mean.

5.2 Scenarios

To distinguish results concerning different models, data sets, and the data set's variants, specific scenario abbreviations are introduced. The two fundamental models are the [**A**]rtificial neural network (Section 5.3) and the [**W**]eighted least squares model (Section 5.1). Furthermore, we differentiate between the [**M**]easured and [**S**]ynthetic data set, where the measured data set is made up of the field measurements and the synthetic data set constitutes the grid simulation. For noise effect investigation, the original [**C**]lean data set is extended by a [**N**]oisy variant. While measured data is inherently noisy, an additional noisy variant for the measurement data is introduced due to the assumption that some of the measurement noise has been cleaned, as discussed in Section 3.1.4. Each scenario is noted with its unique combination of bracketed letters. An overview of the resulting scenarios is depicted in Figure 5.4.

5.3 Model Architecture

When designing an ANN for a specific task, there are different design decisions to be made. One is the fundamental architecture of the model, for example MLP, Graph Neural Network (GNN), Long Short-Term Memory (LSTM), or any variation and combination of those. An overview of different architectural approaches in the literature was given in Subsection 2.3.6. A second design decision is choosing the input and output features based on the available data.

The model chosen for this thesis is a MLP. This thesis aims at investigating the underlying transferability of synthetically trained supervised learning approaches. Since more complex model architectures build on top of the basic MLP components, assessing MLP performance characteristics functions as a baseline. Furthermore, in tests that were run, LSTMs and GNNs, as well as model ensembles including those

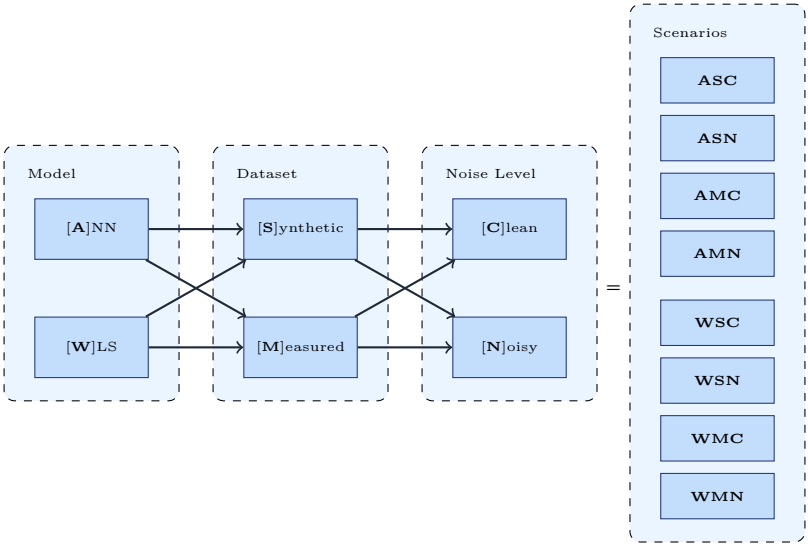


Figure 5.4: Overview of different scenarios, consisting of the model, data set, and noise level.

specialized models and a MLPs did not improve performance significantly. Although more research needs to be done to validate these resutls.

The choice of input and output data, was oriented around the available measurement data, as the overall goal is to assess the ability of a synthetically trained model to be transfered to a real-world measurement environment. This is especially true for the input signals, since otherwise parts of the model's expected input vector would be missing, when testing on measurement data. With respect to output data, any evaluation of the model's performance on measurement data is limited to the available measurement signals. While more outputs can be generated, their validity can only be assessed against the snythetic data set.

Given the available measurements, discussed in Subsection 3.1.2, the input data is chosen to include the voltage magnitude x_v and active power measurement x_p. Since the voltage angle is measured as it's momentary value, it's high frequency and the lack of perfect synchronicity of the measurement devices render the values uncomparable and thereby not usable in the context of the DSSE. The model output is chosen to entail voltage magnitude y_v, voltage angle y_a, and line utilisation y_l, to

monitor voltage, as well as current congestions. For time step t, the input $\vec{x_t}$ and output $\vec{y_t}$ vectors are defined as follows:

$$\begin{aligned} \vec{x_t} &= \begin{bmatrix} \vec{x_{v,t}} \\ \vec{x_{p,t}} \end{bmatrix} \\ \vec{y_t} &= \begin{bmatrix} \vec{y_{v,t}} \\ \vec{y_{a,t}} \\ \vec{y_{l,t}} \end{bmatrix} \end{aligned} \tag{5.4}$$

The DSSE calculation of model M is defined as:

$$M(\vec{x_t}) = \vec{y_t} \tag{5.5}$$

5.4 Model Evaluation

This chapter evaluates the performance of the proposed NSE model. The model's accuracy is compared against a conventional WLS algorithm. Initially, a performance baseline is established using synthetic simulation data. Subsequently, the model's transferability is analysed by evaluating the synthetically trained model on real measurement data.

5.4.1 Noise Scenarios

When training a model on synthetic data, the data is void of any measurement noise that one would expect in its field application. In this subsection, the effect of the introduction of synthetic noise during training is investigated. For that two questions need to be answered. First of all, whether or not the introduction of measurement errors to the model's features during training, improves its ability to minimize noise. And secondly, how the level of measurement error during training and a deviation in training and testing error distributions affects the estimation performance. For the investigation, multiple models are trained with a varying degree of noise levels $0\ \% \leq e_{\%} \leq 4\ \%$. The noise for testing the model is generated with a maximum error of $e_{\%} = 1\ \%$. The specific noise values are sampled according to Equation 3.3. The model output vector comprises measured and unmeasured buses to analyze its noise reduction capabilities on both sets. To assess the performance, the Root Mean Square Error (RMSE) of the model estimation is calculated using the test set. Results are split among measured and unmeasured buses and shown in Figure 5.5.

It becomes apparent, that while the RMSE is lowest when the model is trained on the same noise level as the test set, the difference is relatively small for values below 2 %. Notably, the model with clean samples performs similarly to the model with matching noise at 1 %. The difference between the RMSE of measured and unmeasured buses is very low which signifies that most of the estimation error is attributable to the noise rather than failing to estimate unmeasured nodes. The estimation performance of the model gets significantly worse for $e_{\%} \in \{3, 4\}$. In those cases the model overcorrects for the high error magnitudes it has encountered during training.

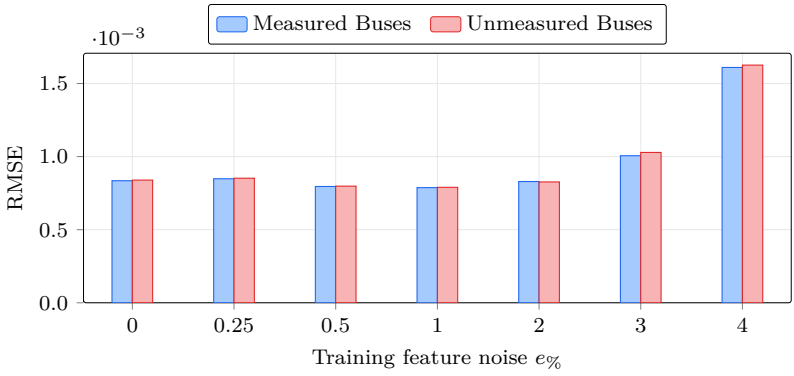


Figure 5.5: RMSE of estimation for varying feature noise during training. Noise during testing is kept constant at 1 %.

To estimate the amount of noise reduction of the model, the ratio between the RMSE of the model prediction $\vec{y_i}$ (RMSE_y) and the RMSE between the noisy model input $\vec{x_i}$ and the true value $\vec{x_{\text{true}}}$ (RSMSE_x) is calculated. The relative noise reduction NR is defined as:

$$\text{NR} = 1 - \frac{\text{RMSE}_y}{\text{RMSE}_x} \tag{5.6}$$

The resulting noise reduction for the aforementioned scenarios is illustrated in Figure 5.6. All scenarios between 0 % and 2 % confirm very similar performances and

show a noise reduction of close to 80 %. The achievable noise reduction gets worse for higher noise percentages.

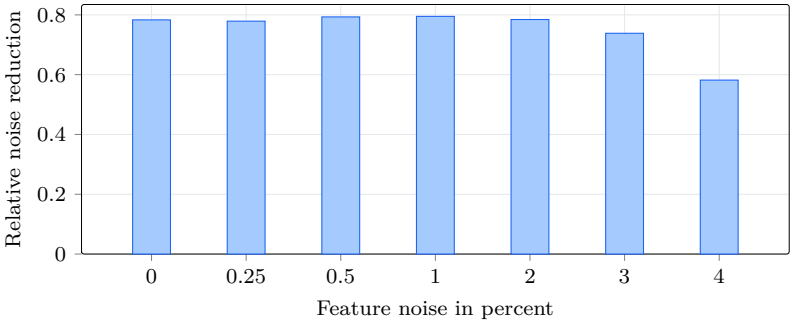


Figure 5.6: Relative noise reduction for varying feature noise during training. Noise during testing is kept constant at 1 %.

Since the model is very effective in noise reduction without being trained on noisy features and there is some uncertainty about the actual noise level in the measurement data, as discussed in Section 3.1.4, no measurement noise is applied to the training data for the NSE model.

5.4.2 Evaluation with Synthetic Data

Assessing the model performance is initially done using the synthetic simulated data set. Since the simulated data set is complete, the state of the unmeasured buses is known and can be used for calculating the estimation deviation. The estimation deviation or error of the estimation for sample t is described as

$$\vec{\Delta_{y,t}} = \vec{y_t} - \vec{y_t}^* \tag{5.7}$$

where $\vec{\Delta_{y,t}}$ represents the concatenated error vector comprising the deviations of voltage magnitude $\vec{\Delta_{v,t}}$, voltage angle $\vec{\Delta_{a,t}}$, and line utilization $\vec{\Delta_{l,t}}$ such that:

$$\vec{\Delta_{y,t}} = \begin{bmatrix} \vec{\Delta_{v,t}} \\ \vec{\Delta_{a,t}} \\ \vec{\Delta_{l,t}} \end{bmatrix} \tag{5.8}$$

Additionally, success criteria are introduced that classify an estimation as successful or unsuccessful based on the deviation magnitude. The criteria are based on [65]. While there are no official requirements for MV-DSSE, the accuracy can be contrasted against the recently suggested DSSE accuracy requirements by the Forum Network Technology and Network Operation (FNN) of the Association for Electrical, Electronic & Information Technologies (VDE). The report suggests a 99^{th} percentile maximum deviation of 2 % for voltage estimation and 10 % for current estimation. Compared to the suggestion the C1 criteria the allowed voltage magnitude deviation is more strict allowing maximum deviations of 1 %. Given the overall tight voltage band of the grid (see Section 3.1.5), a stricter criterion is reasonable. Furthermore, the MV-DSSE is expected to be more accurate due to an overall smaller volatility of the power consumption. Thus, having a second stricter C2 criterion helps evaluate the estimation performance [113].

Table 5.1: Success criteria for state estimation evaluation.

Criterion	$\lvert\Delta_v\rvert$ [%]	$\lvert\Delta_l\rvert$ [%]	$\lvert\Delta_a\rvert$ [ř]
C1	< 1.0	< 10.0	< 0.1
C2	< 0.5	< 5.0	< 0.05

The ANN performance is contrasted against the WLS performance using the model described in Section 5.1. Additionally, models are evaluated using clean and noisy input data. The resulting scenarios are ASC, ASN, WSC, and WSN with reference to the scenario description in Section 5.2. Hyperparameters of the ANN were calculated using hyperparameter tuning. The resulting hyperparameters of the model after hyperparameter tuning are listed in Table 5.2.

Voltage Magnitude Estimation

The distribution of percentage voltage magnitude deviation across all scenarios, as well as the individual success rates, are depicted in Figure 5.7, and Table 5.3 shows

Table 5.2: Resulting MLP hyperparameters after hyperparameter tuning.

Hyperparameter	**Setting**
Number of MLP layers	4
Layer size	150
Activation Function	Leaky ReLU
Loss Function	MSE-Loss
Learning Rate	7.30×10^{-4}
Optimizer	Adam
Batch Size	32

relevant estimation deviation statistics. P_{95} is the 95th percentile. Given a clean synthetic data set, the WLS model outperforms the ANN model. Overall, both models perform very well, scoring 100 % for both success criteria. When analyzing the performance on the noisy data sets, one can see that the ANN is better equipped to handle the synthetic inaccuracies to the WLS model with a mean deviation that is 70 % lower.

Table 5.3: Summary statistics of voltage magnitude estimation deviation using the synthetic data set.

Scenario	**Mean**	P_{95}	**Max**	**C1**	**C2**
ASC	0.0079	0.0205	0.5297	1.00	1.00
WSC	0.0032	0.0151	0.3371	1.00	1.00
ASN	0.0628	0.1574	0.5370	1.00	1.00
WSN	0.2056	0.5223	1.4357	0.99	0.94

Voltage Angle Evaluation

The voltage angle estimation performance is depicted in Figure 5.8 and the corresponding estimation deviation statistics are noted in Table 5.4. In comparison, the ANN achieves much better performance. Interestingly the WLS's estimation for voltage magnitude is much more accurate than the voltage angle estimation. The

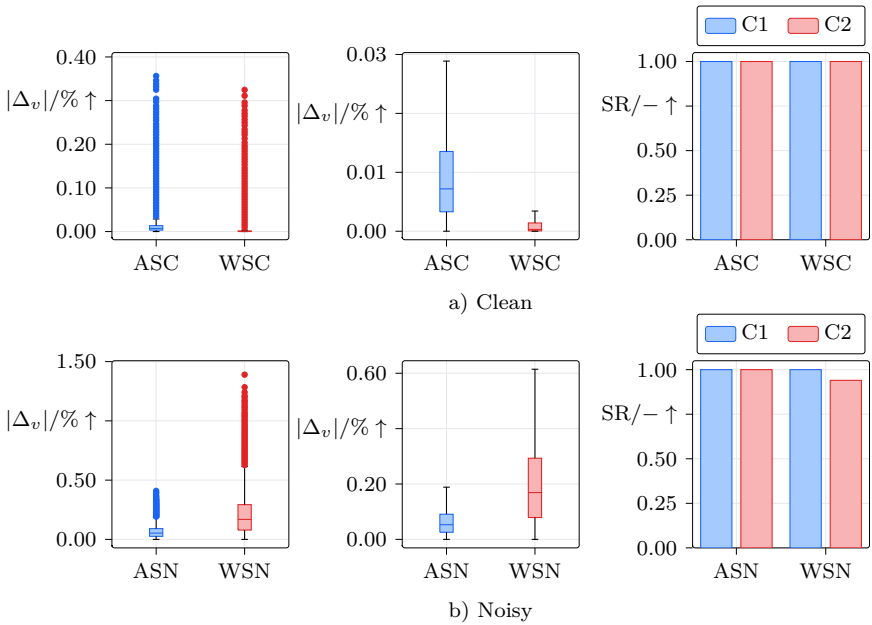


Figure 5.7: Overview of voltage magnitude deviations for the synthetic data set. Subplots show voltage magnitude deviation box plots with (left) and without (middle) outliers, as well as the model's success rate across estimations (right).

WLS model fails a significant amount of estimations with respect to both success criteria in both types of scenarios.

To investigate the performance difference further, the time series with median daily deviation across all nodes of the WLS model is contrasted with the ANN prediction for the same day. Figure 5.9 shows the corresponding time series. It can be seen that the WLS estimation fluctuates around the true values. A reason for that might be a higher convergence criteria, caused by the proportional backoff, which is depicted by the dashed gray line with circular markers. Visually this seems to explain at least some of the deviation. To further investigate the potential correlation, the Pearson correlation between the average angle deviation across all nodes and the convergence criterion for all samples was calculated. The resulting Pearson correlation of 0.44 suggests a moderate linear relationship. Although not a perfect measure, as the

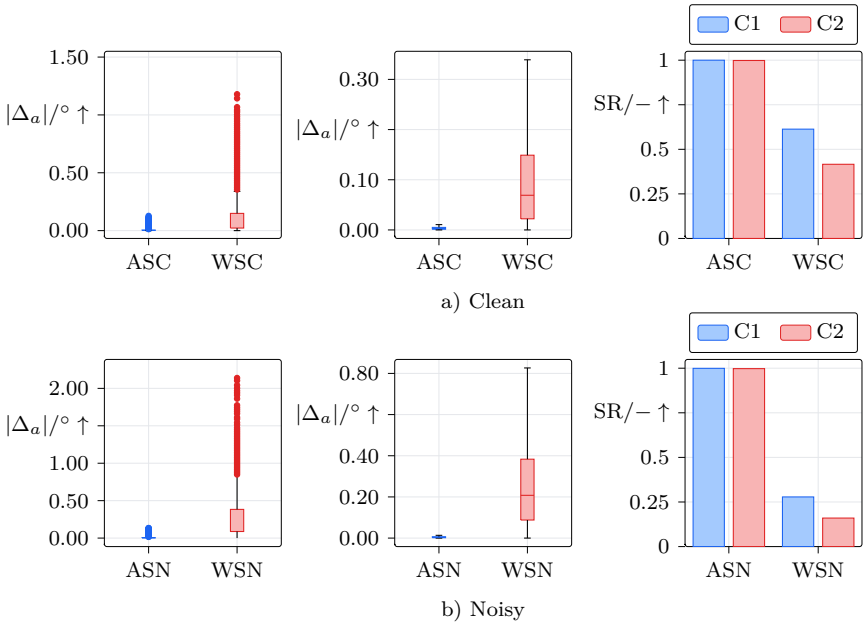


Figure 5.8: Overview of voltage angle deviations for the synthetic data set. Subplots show voltage angle deviation box plots with (left) and without (middle) outliers as well as the models success rate across estimations (right).

Table 5.4: Summary statistics of voltage angle estimation deviation using the synthetic data set.

Model	**Mean**	P_{95}	**Max**	**C1**	**C2**
ASC	0.0037	0.0086	0.1481	1.00	1.00
WSC	0.1078	0.3601	1.3521	0.61	0.42
ASN	0.0049	0.0113	0.1587	1.00	1.00
WSN	0.2645	0.7152	2.2729	0.28	0.16

convergence criterion is updated stepwise and is not a continuous value, it gives some information about the interdependence. The results suggest that the WLS struggles to converge, which loosens the convergence criterion, which leads to higher deviations. While this explains some of the deviations, there are time steps in which the convergence criterion is low that exhibit comparatively high deviations. A possible way to improve WLS results would be to incorporate additional pseudo-measurements or improve their accuracy.

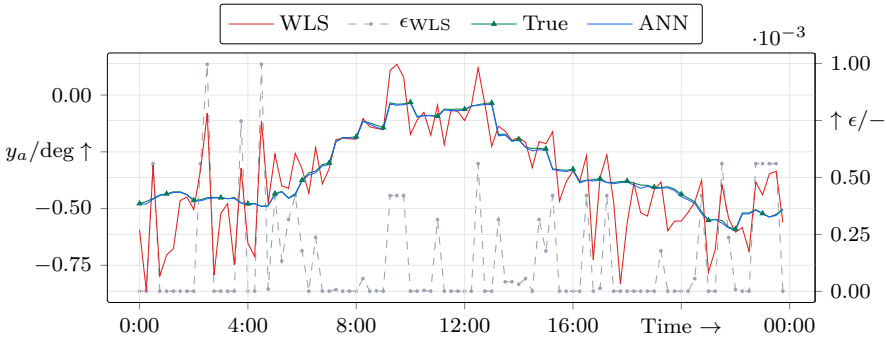


Figure 5.9: Exemplary day with a median average deviation of the voltage angle estimation of the WLS model, with ANN model for reference.

Line Utilization Evaluation

Finally, both model's line utilization estimation performance is shown in Figure 5.10 with summary statistics noted in Table 5.5. Similarly to the voltage angle estimation, the ANN demonstrates superior performance.

In Figure 5.11, the time series with median average deviation for the line utilization is depicted alongside the convergence criterion. We can see stark deviations between the true line utilization and the model estimation in certain time steps. Differing from the voltage angle estimation, there seems to be no correlation between the estimation error and the convergence criteria. This is confirmed by a Pearson correlation of -0.22 across all estimations. Since the WLS only actively estimates voltage magnitude, the line utilization is determined physically given this system state. As

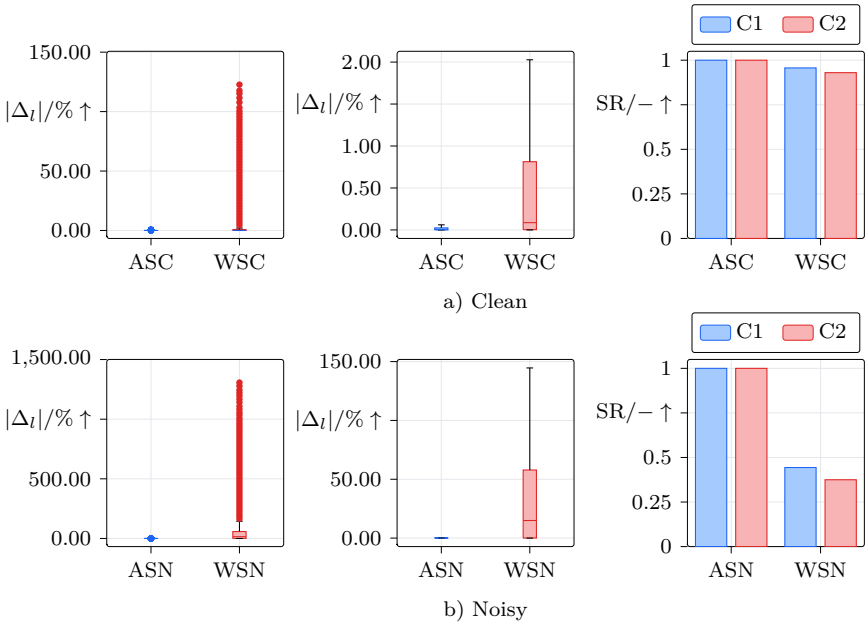


Figure 5.10: Overview of line utilization deviations for the synthetic data set. Subplots show line utilization deviation box plots with (left) and without (middle) outliers, as well as the models success rate across estimations (right)

Table 5.5: Summary statistics of model line utilization estimation deviation.

Model	**Mean**	P_{95}	**Max**	**C1**	**C2**
ASC	0.0208	0.0941	0.6751	1.0000	1.0000
WSC	1.8906	8.2639	134.0342	0.9566	0.9300
ASN	0.0216	0.0909	0.9282	1.0000	1.0000
WSN	53.8395	252.0738	1363.7538	0.4436	0.3750

slight voltage magnitude and angle deviations can infer drastically different current flows, higher line utilization deviations can be expected.

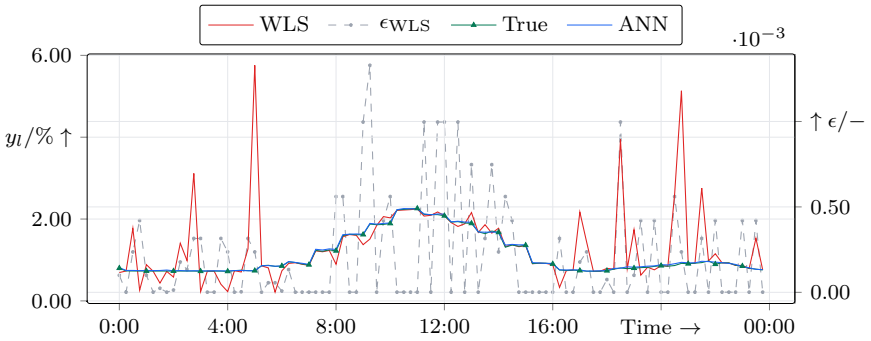


Figure 5.11: Exemplary day with a median average deviation of the line utilization estimation of the WLS model, with ANN model for reference.

Summary

The synthetic data evaluation demonstrates that both the ANN and WLS models achieve good performance in estimating voltage magnitudes. The WLS outperformed the ANN model on the noiseless data set. However, the deviations of both models were sufficiently small that the performance difference has negligible practical impact. When adding noise, the ANN model has shown improved capabilities to reduce noise and overall better performance than the WLS model. For voltage angle and line utilization, estimation the ANN has shown very good performance, reaching 100 % across both success criteria regardless of noise. WLS performance overall was unsatisfactory, failing with respect to both success criteria a significant amount of times.

5.4.3 Evaluation with Measured Data

While approaches in the literature stop and deliver their assessment after checking synthetic performance, this does not evaluate the model under real operating conditions. The synthetically trained model must prove transferability to these op-

erating conditions, which most importantly means the transfer from synthetical to measurement data as this changes the underlying data distribution.

Round-Robin Training and Evaluation

One central challenge given the problem definition is the model performance evaluation. Since measurements from monitored buses serve as input and estimated values for unmonitored buses serve as output, direct validation of predictions remains impossible due to missing ground truth measurements at unmonitored locations. Therefore, an estimation assessment is not possible. To circumvent this problem, a round-robin training and evaluation approach is suggested.

The basic concept of round-robin training and evaluation involves sequentially excluding a measured node from the training data set while adding it to the set of nodes to be estimated. For each configuration, the model trains with synthetic data then tests against measurement data. Predictions for the measured node that was hidden during training can be compared to actual measurements to assess model performance. Repeating this procedure for each measured node creates a comprehensive data set with predictions for all measured buses, enabling thorough model performance assessment. This approach assumes performance at measured buses indicates performance at unmeasured buses throughout the grid. Figure 5.12 illustrates this procedure. The previously calculated hyperparameter settings are used to parameterize the model (see Table 5.2).

Overall Evaluation

Before evaluation, the measurement data set is preprocessed as discussed in Section 3.1.4. Furthermore, the previously trained min-max scalers are applied to normalize the values between zero and one. Unfortunately, only voltage magnitude predictions can be assessed, as line utilization measurements are unavailable, and the measured angles are unsynchronized absolute values.

Figure 5.13 depicts the overall deviation distribution and the C1 and C2 success rates for both models on measured data. Using the clean measurements depicted in Figure 5.13a, the WLS model outperforms the ANN model slightly. While the WLS only fails with respect to the C2 criterion on a few outlier estimations, the ANN model achieves around 90 %. The performance evens out when adding synthetic noise to the measurement data, as seen in Figure 5.13. The general error distribution between both models is very similar, with the ANN model having more outliers. Due

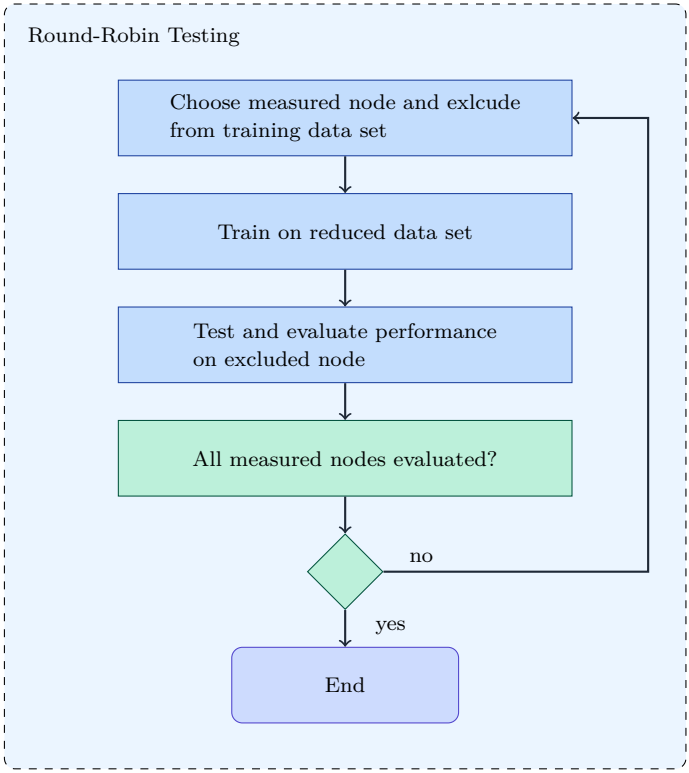


Figure 5.12: Round-robin training and evaluation to gauge model performance on measurements.

to the outliers, slightly worse performance across both criteria can be seen, while both models still achieve upwards of 90 % across both criteria and close to 100 % for C1.

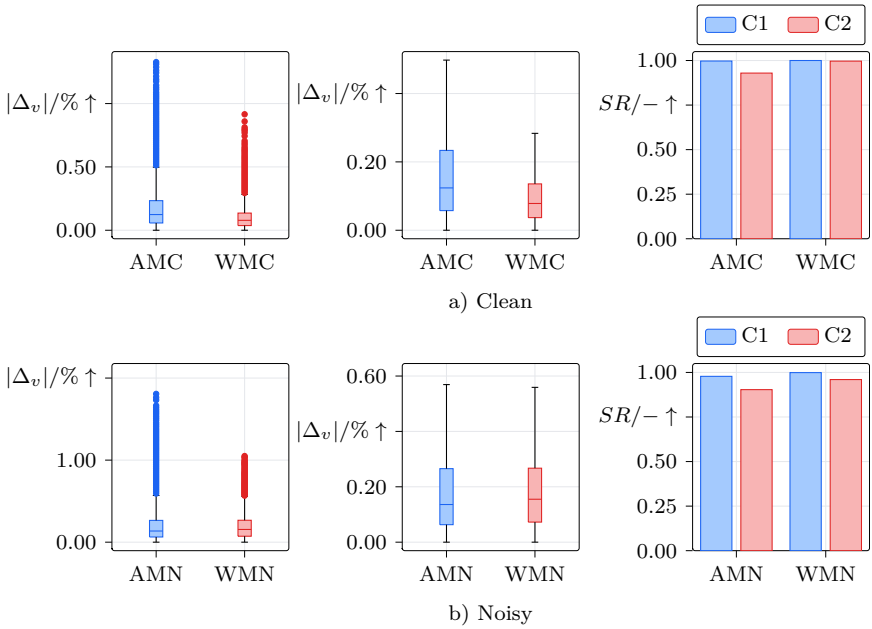


Figure 5.13: Overview of voltage magnitude deviations for the measured data set. Subplots show voltage magnitude deviation box plot with (left) and without (middle) outliers, as well as the model's success rate across estimations (right).

To get a sense of the actual predictions of both models, Figure 5.14 shows the true value and predictions with different average daily deviations. The subplots a), b), and c) show the maximum, median, and minimum daily average deviation time series. While the minimum and median deviation cases illustrate comparably good performance of both models, we can see an offset in the maximum deviation case.

Overall, the average deviation of both models increased drastically, illustrating that performance on a synthetic data set is not necessarily indicative of performance

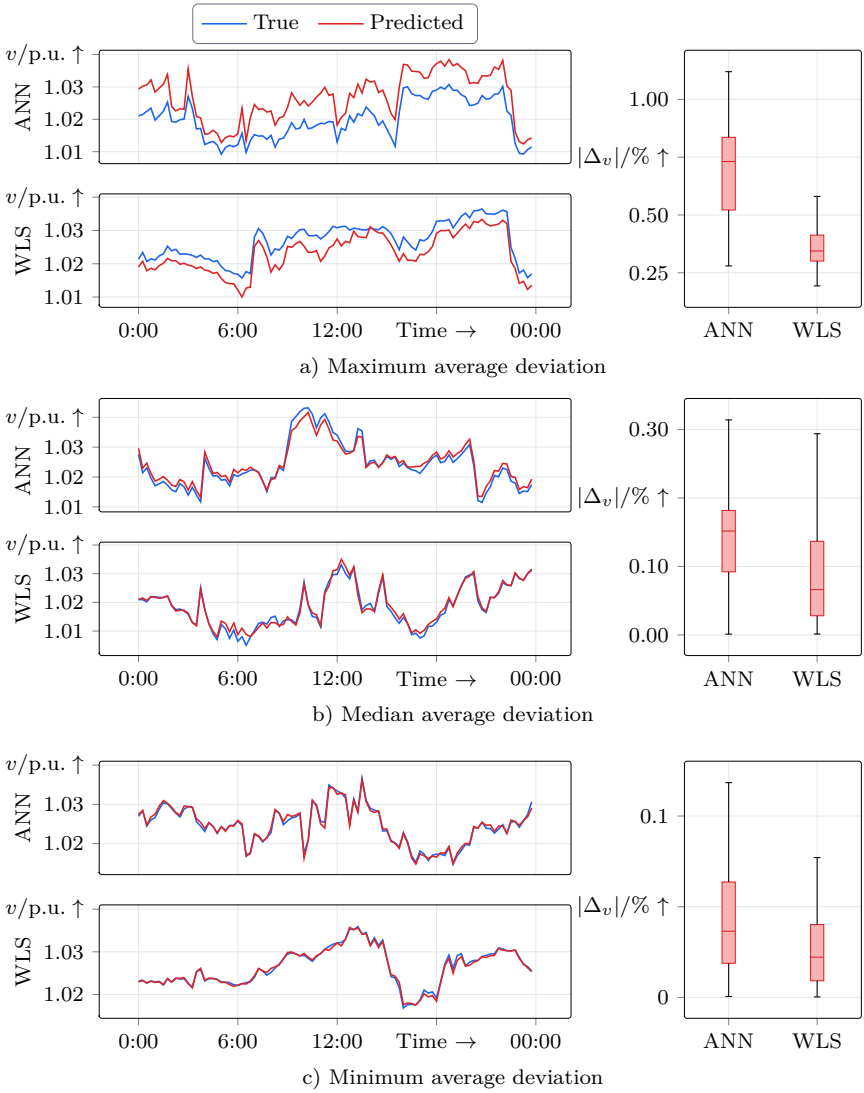


Figure 5.14: Maximum, median, and minimum average daily deviation on clean measurement data for ANN (AMC) and WLS (WMC) model

on field measurements. It has to be noted that evaluation on field measurements presents a challenge as some of the observed model 'error' might actually be attributable to measurement inaccuracies rather than model deficiencies. The WLS model performed well on voltage magnitude estimation on the measured data set, suggesting that the implemented pseudomeasurement strategy is effective for voltage magnitude estimation in real-world conditions. Although the neural network model demonstrates adequate performance overall, it does not improve upon the WLS performance. This discrepancy between synthetic and real-world performance highlights the challenges of transferring models trained on simulated data to actual field conditions. The ANN model's inability to outperform the conventional WLS approach on measured data suggests that either the synthetic training data does not fully capture the complexities of real-world grid behavior, or that the neural network architecture requires further refinement to better generalize to measurement data distributions.

5.4.4 Bus-Specific Evaluation

While the previous investigations measured the model's overall performance, the following will focus on performance estimation across the different buses of the grid on the measurement data set. Figure 5.15 shows the success rates with respect to individual buses. It can be seen that there is a significant variation in estimation performance for the ANN model. While approximately half of all buses achieve 100 % or close to 100 %, the worst-performing buses have a success rate around 70 %.

Model training is not deterministic. Due to different initialization values of the neuron weights and shuffling of the input samples, the resulting model weights differ across different training runs, leading to different model estimations. Twenty full iterations of the round-robin training and testing were performed to analyze variations across training runs. The different estimation results for C2 performance and differences in estimation deviation are plotted in Figure 5.16a and Figure 5.16b, respectively. The buses have been sorted by maximum C2 count difference across samples. It becomes evident that there is a noticeable variation in estimation performance throughout all buses. This effect is especially pronounced for a subset of the buses, with a mean deviation for the same bus that can be almost twice as high for the same bus on different training runs and an absolute C2 rate difference close to 30 %. For different training runs, the resulting weights of the ANN can be different. While the resulting performance on the synthetic validation and test

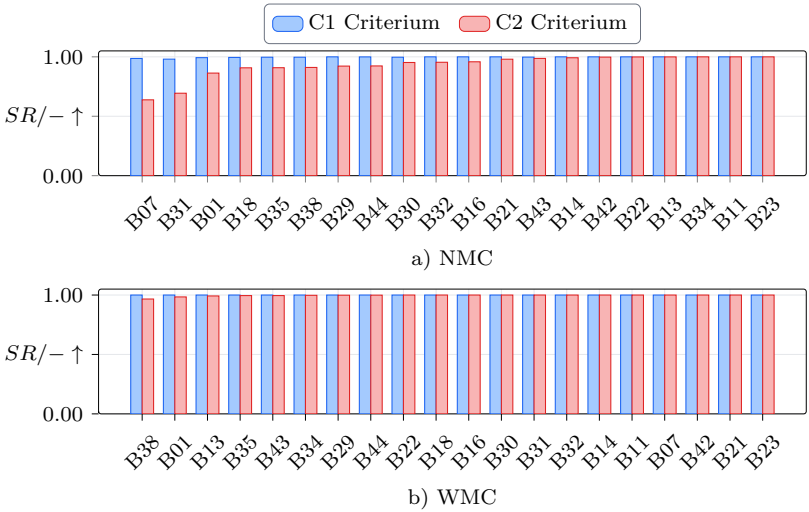


Figure 5.15: Bus-specific C1 and C2 success criterion for voltage estimation.

data is very small this is not true for the performance on the measurement data. This discrepancy suggests a difference between the synthetic and measurement data distributions. The neural network appears to find multiple local minima during training that perform similarly well on synthetic data but generalize differently to real measurements.

Figure 5.16 visualizes the position of the four nodes with the highest performance variation. One would assume that performance decreases with distance to the nearest measured buses. Looking at bus B44, this assumption does not hold up, given that the estimation for the bus is among the worst-performing buses while being directly connected to two measured buses where estimation accuracy is expected to be very good.

Figure 5.18 shows the bus positions in more detail. Bus B44 is directly connected to B42, which is connected to B32. In the model's estimation, much of the error stems from a prediction lower than the actual voltage.

A typical example is shown in Figure 5.19. As shown in Figure 5.19b, the active power of all three buses is positive throughout all time steps. The current direction indicates a voltage drop from B32 to B42 to B44. The voltages in Figure 5.19a

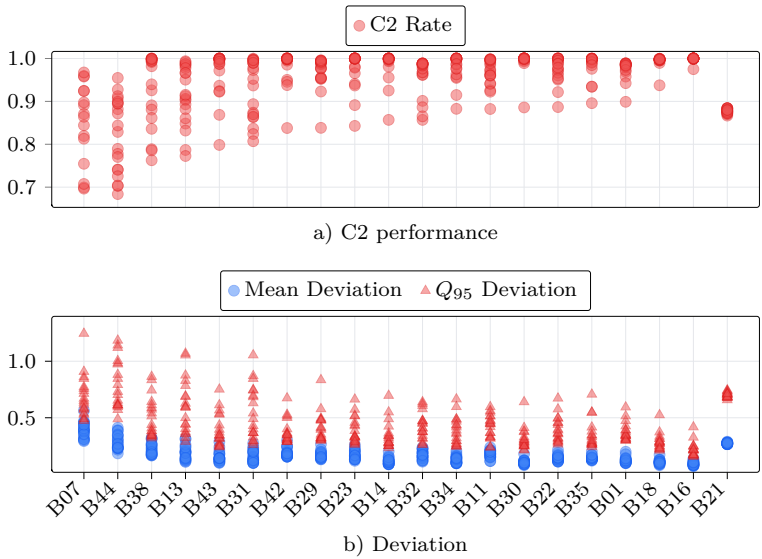


Figure 5.16: Mean and 95th quartile voltage deviation on measurements of different training iterations.

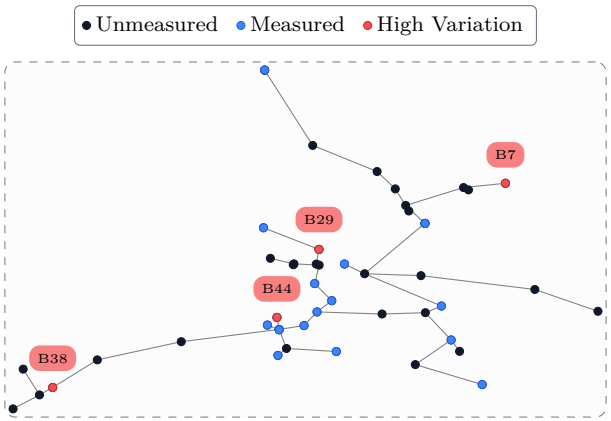


Figure 5.17: Grid graph plot with marked nodes that contain the highest variation.

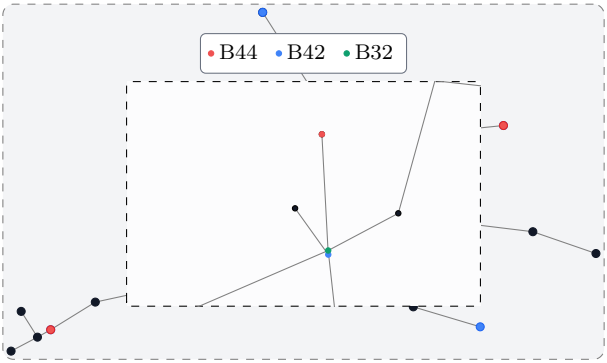


Figure 5.18: Grid graph plot with marked nodes that contain the highest variation.

do not reflect this assumption as the voltage of B44 specifically is greater than the voltages of the other buses throughout most of the time steps. A qualitative assessment of the B44 model prediction shows that it better reflects this assumption. This illustrates one of the central challenges when dealing with real measurement data. Since we do not know the actual ground truth of the data points, we can only take educated guesses when encountering anomalies. In this case the issue seems to be the measurement. Given that the values are measured on the low voltage side of the MV/LV transformers, a reason could be a mismatch between the transformer's real and assumed turns ratio.

In summary, the bus-specific evaluation revealed significant ANN performance variations across buses for different training runs. While multiple training outcomes perform similarly on synthetic data, they generalize differently to real-world measurement data. Analysis of specific buses, like B44, revealed instances where the model's prediction seemed more physically plausible than the actual measurement. These discrepancies highlight the challenge in precise performance evaluations given inaccurate measurement data or an imprecise grid model parameterization.

5.4.5 Computation Time

One of the advantages of many of the ML models is their superior computation times. Most of the calculatory effort is shifted from the model calculation to the model

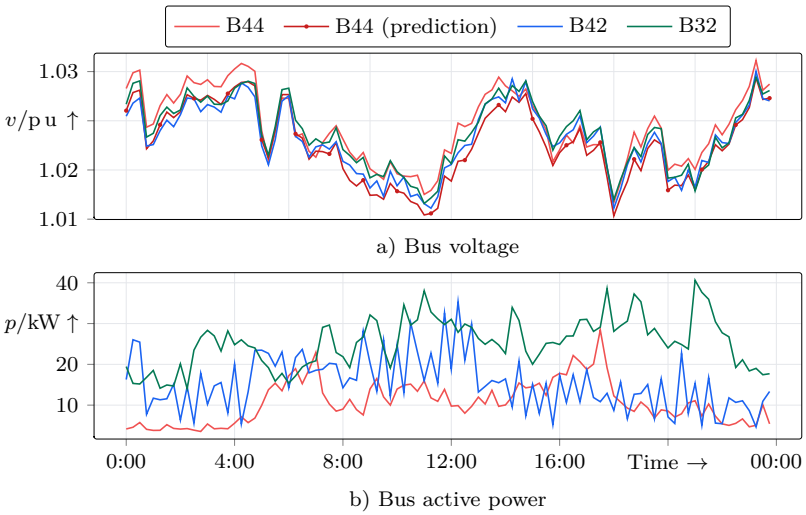


Figure 5.19: Measurements and prediction for bus B44 and surrounding buses.

training. As such, the model inference is often magnitudes faster than conventional methods, especially if those involve optimization procedures.

Multiple experiments were run to analyze differences in computation speeds between the NSE and WLS models. Since the inference of a MLP comes down to matrix multiplication, the exact input does not factor into the overall computation time. For the WLS, they do, as the model can converge faster, given more precise measurements. Hence, a comparison of the computational speed between the ASC scenario and all WLS scenarios was performed. Given that proportional back off is used in the WLS scenario if the calculation does not converge, a differentiation is made between the time for a regular run and the time of all runs of the algorithm until convergence. The latter scenario is noted by an additional BO suffix, for example, WSC-BO.

The experiments were run on an M1 Macbook Pro with 16GB of RAM and an M1 Pro chip with eight performance and two efficiency cores. It is important to note that this comparison does not only compare the theoretical algorithm speed but also implicitly its specific implementation. So, while tremendous work has been done to speed up ML model inference, the same can not be said for the specific WLS al-

gorithm. Therefore, performance gains by improving the algorithm implementation and better usage of concurrency can be expected.

Figure 5.20 depicts the duration of the different scenarios in seconds, and a summary of the runtime statistics is shown in Table 5.6. A noticeable disparity between the three algorithm groups can be observed. The neural state estimation exhibits runtimes about three orders of magnitude smaller than those of the fastest WLS scenario. A significant increase in runtime can be observed when comparing the single iteration WLS time versus the complete runtime, including the backoff. The mean iteration of the proportional backoff scenarios lie between roughly 10 and 30, issueing a 10 to 30 times slowdown. The iteration amount depends on the backoff factor chosen for the convergence. Across both WLS scenario groups, it becomes visible that the introduction of noise significantly slows down the algorithm because more iterations are needed to achieve the convergence criterion. This is especially apparent when comparing the synthetic backoff scenarios WSC-BO and WSN-BO. The runtimes of the measurement scenarios are similar to the WSN-BO scenario, with a slight increase when additional noise is added.

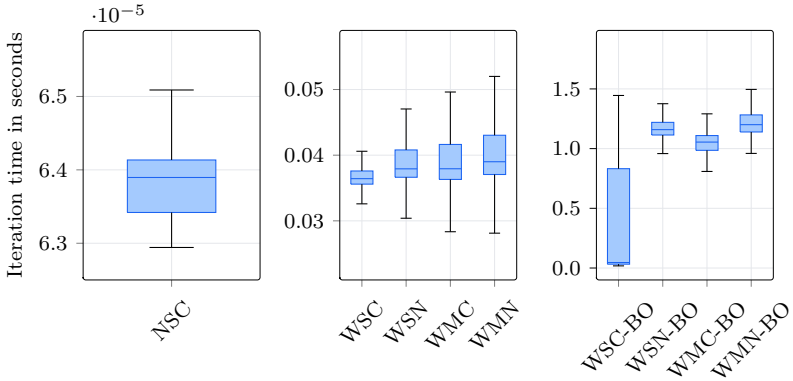


Figure 5.20: Comparison of compuational time of the different scenarios. Outliers have been excluded for better readability.

Table 5.6: Computation time statistics for the different scenarios.

Algorithm	**Min** [s]	**Mean** [s]	**Max** [s]	**Standard Deviation** -	**Mean Iterations** -
ASC	6.00e-5	6.60e-5	5.55e-3	5.20e-5	-
WSC	0.017	0.038	0.543	0.011	-
WSN	0.017	0.042	0.589	0.016	-
WMC	0.017	0.045	0.559	0.015	-
WMN	0.017	0.043	0.569	0.018	-
WSC-BO	0.018	0.406	1.452	0.402	10.80
WSN-BO	0.035	1.172	1.785	0.108	27.93
WMC-BO	0.021	1.113	2.994	0.140	24.58
WMN-BO	0.841	1.233	2.617	0.155	28.40

5.5 Reduced Observability

Previous investigations evalued the algorithms on one specific measurement configuration and by extension one specific observability rate. To investigate the effect varying observability rates have on estimation performance of the different algorithms, additional scenarios with a reduced set of measurements are investigated.

5.5.1 Reduction Algorithm

The overall goal of the reduction algorithm is to reduce the grid observability by removing the most insignificant measurements first. Given a fixed number of measurements more information is gained from a sparse but even distribution of measurements thorughout the grid as opposed to a dense measurement distribution in a specific area of the gird. Thus, the measurements are removed in order of their least distance to the next measurement. Given an initial observability of around 42 % for each scenario, an observability rate $o_i \in \{0.3, 0.2, 0.1, 0.05\}$ is defined. Given the observability, the amount of buses to remove equals

$$r = \lfloor o_i \cdot |B_{\text{measured}}| \rfloor \tag{5.9}$$

For each bus to remove, the shortest path distance between all measured bus pairs b_i and b_j where $b_i, b_j \in B_{\text{measured}}$ is searched.

$$(b_i, b_j) = \underset{b_i, b_j \in B_{\text{measured}}}{\operatorname{argmin}} \quad d(b_i, b_j) \tag{5.10}$$

From this bus pair, we find the node with greater distance to the slack node b_{slack}

$$b_{\text{remove}} = \begin{cases} b_i & \text{if } d(b_i, b_{\text{slack}}) > d(b_j, b_{\text{slack}}) \\ b_j & \text{otherwise} \end{cases} \tag{5.11}$$

Bus b_{remove} is subsequently removed from the set B_{measured} and added to B_{removed}. The process is repeated until $|B_{\text{removed}}| = r$. Figure 5.21 summarizes the process.

To illustrate the results of the observability reduction, Figure 5.22 represents the graph model given an observability rate of 10 %. The algorithm successfully reduces close measurements, leaving the remaining measurements spread out.

5.5.2 Evaluation

Given the reduced measured buses for observability rates $o_i \in \{0.3, 0.2, 0.1, 0.05\}$ the ANN and WLS model are evaluated. Rather than utilizing the round-robin evaluation approach discussed in Section 5.4.3 the model performance was established using the measured but hidden buses. Although this compromises some degree of exactness, this reduces the calculatory expense significantly. Since the approach is taken for the ANN and WLS models and the difference in performance between those models is most important, this approach is deemed sufficient.

Box plots of the resulting voltage magnitude deviations accross all buses and samples for the different observability rates are depicted in Figure 5.23a and the corresponding C2 scores in Figure 5.23b. Subplot (a) shows that while both models experience increased voltage magnitude deviations as observability decreases from 30 % to 5 %, the ANN model consistently maintains lower median deviations and less variability (smaller IQR) compared to the WLS model (red). This difference becomes particularly evident at 10 % and 5 % observability, where the WLS error distribution widens considerably. Subplot (b) complements this by showing the C2 success rate. The NN model retains a high success rate (above 0.9 at 10 %, about 0.89 at 5 %), whereas the WLS model's success rate drops more sharply, falling below 0.8 at 10 %

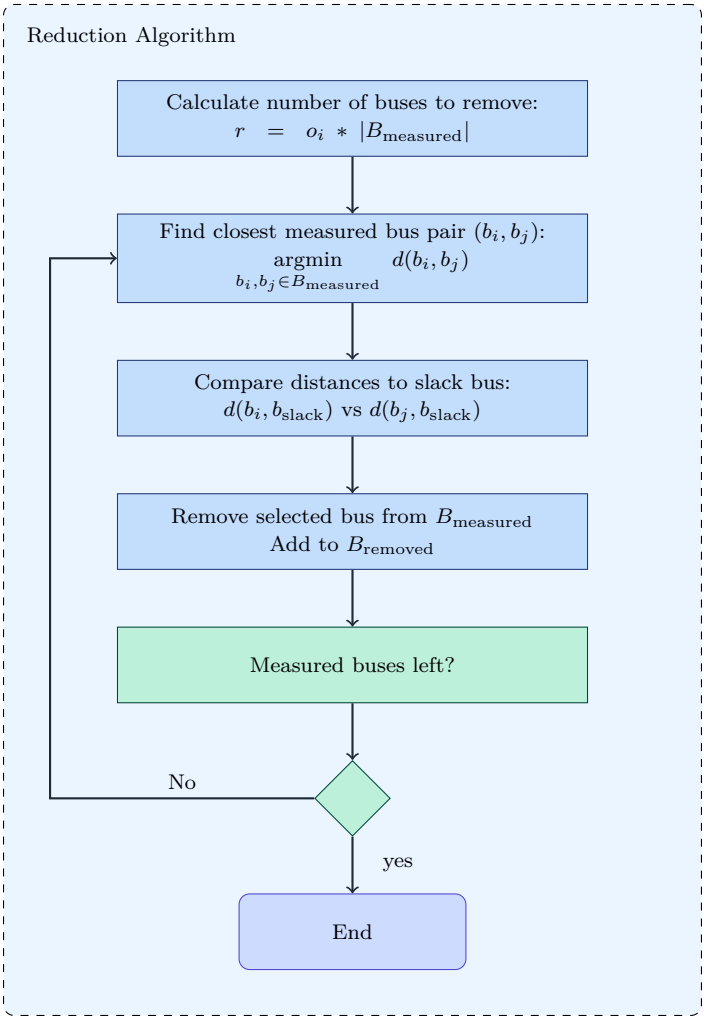


Figure 5.21: Reduction algorithm flowchart.

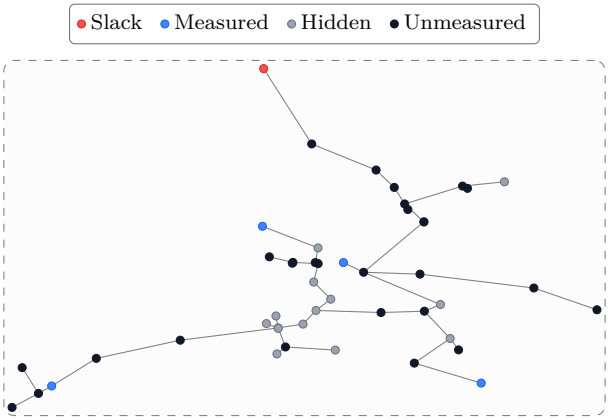


Figure 5.22: Representation of the graph with reduced observability for an observability rate of 10 %.

and approaching 0.6 at 5 %. This demonstrates the NN's superior robustness in providing accurate estimations when measurement data is sparse.

5.6 Robustness against Utilization Changes

Training a model on a static data set, called offline training, results in a static model, operated in a dynamic environment. The environment can be considered dynamic, as its operation states morph over time. One can differentiate between two types of changes in the grid environment, which are topological changes and utilization changes. Topological changes consist of modifications to the grid topology, such as adding new buses and line connections. These changes usually require retraining of the model. Utilization changes involve adding or removing load or generation units at existing buses. An example could be the installation of a new PV unit at a specific bus. This chapter investigates the robustness of the model against such utilization changes.

Two additional data sets are created for the following analysis by changing the system participant models and running the corresponding simulations. The first is the load scenario D_{load} that represents an increase in power consumption within

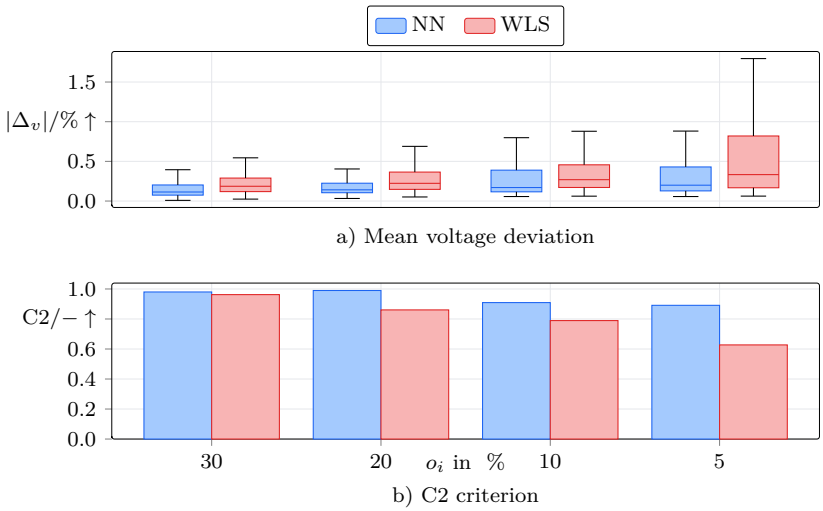


Figure 5.23: Voltage magnitude deviation and corresponding C2 success rate for different observability rates.

the grid. Here each household load consumption was scaled by a factor of two. The second is the generation scenario $D_{\text{generation}}$, where all PV units were scaled by a factor of two. Exemplary resulting bus voltage and power distributions are depicted in Figure 5.24. Due to the shifted load and generation, a slight shift in voltage magnitudes can be observed. Since more load than generation is present in the original grid, the shift of the load shift of the active power and the resulting negative shift in voltage magnitude is more pronounced.

To assess the model performance of the generated scenarios, the scenario data was scaled using the scalers trained during model training. This way, the data is preprocessed like actual measurement data would be preprocessed in a field application. The distributions of the data seen in Figure 5.24 overlap but extend to include higher voltages in the scaled generation scenario and lower voltages in the scaled load scenario. It can be assumed that the model will perform better in the average cases with respect to voltage magnitudes than in the more extreme scenarios that haven't occurred during training. To investigate this assumption, the model was tested on different subsets of the scenarios. In the base scenario, the model deviation is evaluated on the entire synthetic data set. Furthermore, the estimation

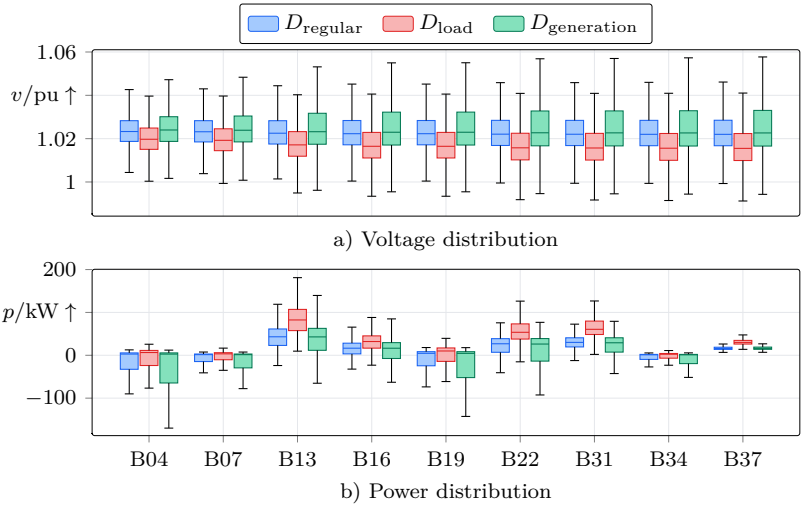


Figure 5.24: Nodal voltage and power distributions for the scaled load and generation scenarios.

error is calculated on the more extreme subsets of the data. The extreme samples are defined as the samples with the highest or lowest average voltages, depending on the scenario. Given all voltage magnitudes $v_{i,t}$ for node i at time t, the average voltage across all nodes is calculated as:

$$\bar{v}_t = \frac{1}{n} \sum_{i=1}^{n} v_{i,t} \tag{5.12}$$

where n is the total number of nodes in the network. Let $0 \leq \alpha \leq 1$ denote the extreme sample fraction, which is defined as the relative amount of considered samples. For each α, a subset $D_{\alpha,\text{generation}}$ is defined containing the samples with the highest average voltages:

$$D_{\alpha,\text{generation}} = \{x \in D_{\text{generation}} \mid \bar{v}_t(x) \geq v_{\text{threshold,generation}}(\alpha)\} \tag{5.13}$$

where $v_{\text{threshold}}(\alpha)$ is chosen such that $|D_{\alpha,\text{generation}}| = \alpha T$, with T being the total number of time steps in the dataset. Similarly, for the load case $D_{\alpha,\text{load}}$ is calculated as follows

$$D_{\alpha,\text{load}} = \{x \in D_{\text{load}} \mid \bar{v}_t(x) \leq v_{\text{threshold,load}}(\alpha)\} \tag{5.14}$$

A box plot of the model estimation deviation distribution for different extreme sample fractions α is shown in Figure 5.25. The model was trained on D_{regular} and all data sets were limited to the data set fraction reserved for testing. Outliers have been omitted for better readability. The procedure for calculating the different D_{regular} subsets mirrors the procedure for $D_{\text{generation}}$ (see Equation 5.13) and D_{load} (see Equation 5.14) respectively. Accordingly, the upper box plot for $\alpha = 1\,\%$ includes the 1 % samples with the highest voltages for $D_{\text{generation}}$ and D_{regular}. The lower box plot for $\alpha = 1\,\%$ includes the 1 % samples with the lowest voltages for D_{load} and D_{regular}. A stark increase in the error rate of the model estimation can be seen for higher values of α. If the model has not encountered the extreme samples as part of its training set, its estimation performance is worse. On average, the model tends to overestimate voltages for $D_{\text{generation}}$ and underestimates voltages for D_{load}. This is notable as one might expect the model to favor prediction ranges of the training data set and thus underestimate the higher voltages of $D_{\text{generation}}$ and overestimate the lower voltages of D_{load}.

5.7 Fine-Tuning

As shown in the last section, the model struggles to estimate values it has not seen as part of its training. A possible strategy to deal with that would be to extend the regular training of the model with a fine-tuning phase. During the fine-tuning, more extreme samples are fed to the model to extend its capabilities.

A common issue with fine-tuning a model to specific parts of a data distribution is the risk of overfitting to those examples. To test this, two fine-tuning data sets $D_{5\%,\text{ generation}}$ and $D_{5\%,\text{ load}}$ were generated. Each consists of the 5 % most extreme samples according to Equation 5.13 and Equation 5.14 respectively. The resulting distribution of those subsets are depicted in Figure 5.26.

The data was divided into 80% training, 10% validation, and 10% test samples. Two models, $M_{\text{generation}}$ and M_{load}, were trained on the regular synthetic data set and then fine-tuned. One was fine-tuned using the $D_{5\,\%,\text{generation}}$ training samples

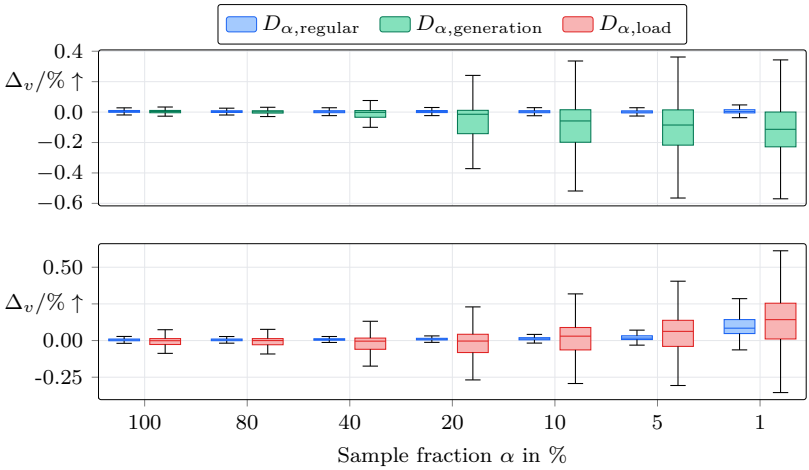


Figure 5.25: Voltage magnitude estimation error for different subsets of the scaled load and generation scenarios.

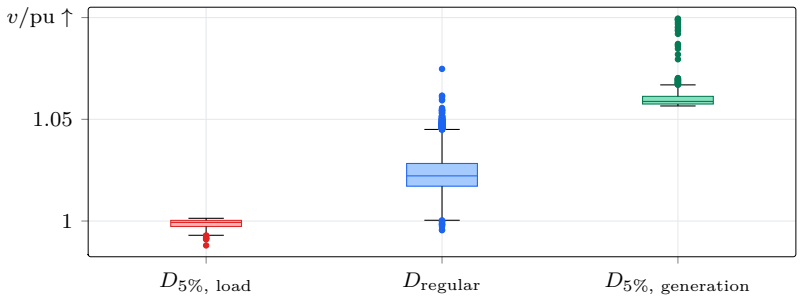


Figure 5.26: Fine-tuning distribution of scaled training data sets.

and one using the $D_{5\ \%,\text{load}}$ training samples. For the training early stopping was utilized, choosing the model that performs best on the respective validation set. The models were then tested on the respective test sets as well as the test set of the original data. Furthermore, the model solely trained on the original data, M_{regular}, was tested on the extreme samples test sets. The resulting average error of the different models on the different test sets is depicted in Figure 5.27. The fine-tuned models $M_{\text{generation}}$ and M_{load} reduce the average absolute estimation error of M_{regular} on $D_{5\%,\ \text{generation}}$ and $D_{5\%,\ \text{load}}$ by 99 % and 90 % respectively. However, the performance of both fine-tuned models on D_{regular} degrades significantly.

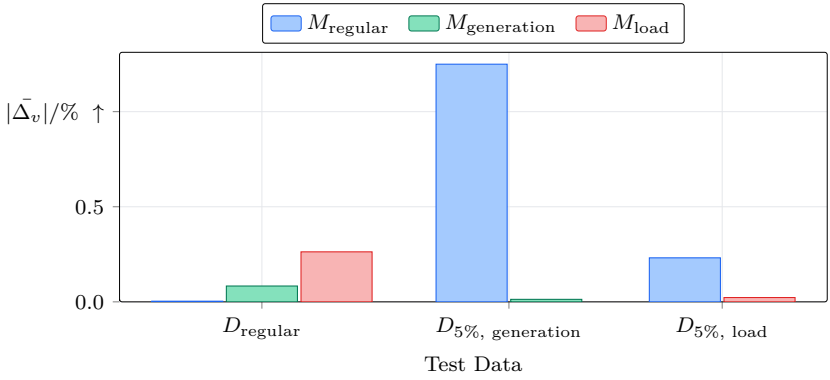


Figure 5.27: Average absolute voltage deviations of models on different test sets. M_{load} and $M_{\text{generation}}$ are fine-tuned on D_{load} and $D_{\text{generation}}$ samples respectively.

To ensure that the models retain their performance on the original data distribution, the fine-tuning training samples are mixed with training samples of the original data set. Figure 5.28 visualizes the general idea.

While the depicted example shows an even distribution of the extreme and the regular samples, different compositions can be used. In order to assess the effect of different fine-tuning compositions, training runs with a varying amount of extreme samples are run and evaluated. The amount of extreme samples is evenly split between $D_{5\%,\ \text{generation}}$ and $D_{5\%,\ \text{load}}$. Performance is measured by evaluating the trained models on the test sets of D_{regular}, $D_{5\%,\ \text{generation}}$ and $D_{5\%,\ \text{load}}$. Figure 5.29

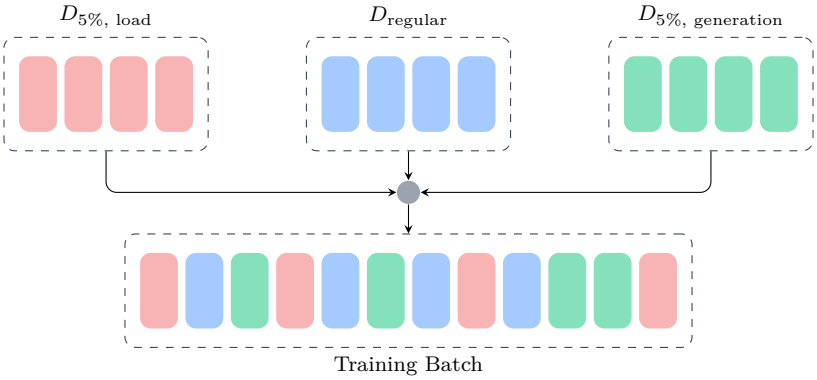


Figure 5.28: Mixed fine-tuning batch creation using samples from $D_{5\%,\ \text{load}}$, $D_{5\%,\ \text{generation}}$ and D_{regular}.

shows the corresponding results. Most notably, introducing just two fine-tuning samples into the data set leads to a dramatic reduction in error on the fine-tuning test sets. Adding four samples still shows a significant improvement in fine-tuning performance with only a limited increase in the error on regular samples. Afterwards, there is a roughly linear tradeoff between performance on the regular test set and the fine-tuning test sets. Furthermore, a sharp increase in the error of the regular test set can be seen when no regular data is present. Notably, only two regular samples are sufficient to keep the error on regular samples relatively low. A tradeoff between specificity and generality of the model can be seen. Widening the training distribution by fine-tuning on more extreme samples leads to a better performance on a wider set of data points, which is an increase in generality of the model. At the same time, the performance on the specific original data distribution gets slightly worse, indicating a loss of specificity.

Overall, fine-tuning the model is an effective way to increase estimation performance on parts of the distribution that were not present in the training data. Including samples from the original distribution prevents a performance degradation on the original data set. The relation of regular to fine-tuning samples leads to trade-off between performance on the regular versus the fine-tuning data. However, this trade-off effect is relatively minor when compared to the substantial performance gains achieved by including the first few fine-tuning samples or regular samples respectively.

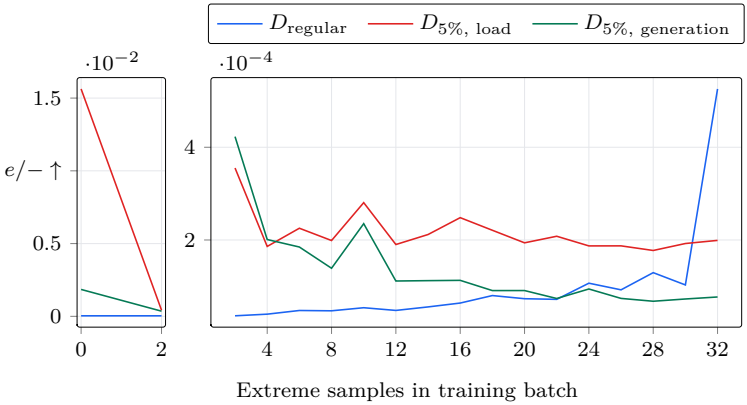


Figure 5.29: Fine-tuned model error with respect to different quotas of fine-tuning samples in 32-sample training batch.

5.8 Uncertainty Estimation

A common concern that hinders the adoption of ML methods, especially in the context of critical infrastructure, is the model's limited transparency. This refers to the difficulty of understanding and interpreting how these models arrive at their predictions. Given that the DSSE is likely to form the basis for operational decision and monitors the health of the electrical components, this concern is well founded. While Chapter 4 discussed methods to check and improve the incoming data for the NSE model, we can not be certain that every error was detected and successfully corrected. Furthermore, the model might encounter out-of-distribution data, which it may not be able to generalize to. To improve the model's transparency, the usage of MC Dropout for uncertainty estimation is proposed and investigated in the following Section. Extending on the MC Dropout estimation, a probabilistic state estimation approach is developed that computes estimation ranges instead of uncertain point predictions.

5.8.1 Monte Carlo Dropout for Neural State Estimation

Dropout is a widely used regularization technique in neural networks, introduced to mitigate overfitting. During training a fraction of neurons is deactivated. This

stochastic deactivation prevents neurons from developing strong co-dependencies, thereby enhancing the model's generalization capabilities. Originally, dropout was designed to be exclusively activated in the training phase, during which each forward pass utilizes a different subset of the model's neurons to improve model training. MC Dropout extends the concept of dropout from the training phase to the testing phase. This approach was first proposed by Gal and Ghahramani in [114]. By applying dropout during both training and testing, multiple forward passes through the network with different subsets of activated neurons can be performed. This process effectively creates an ensemble of models within the same network architecture. Each pass activates a different set of neurons, resulting in slightly different predictions each time. By aggregating these predictions, one can gain insight into the distribution of the network's outputs. The variance or spread of these predictions can then be interpreted as a measure of uncertainty. The procedure is depicted in Figure 5.30.

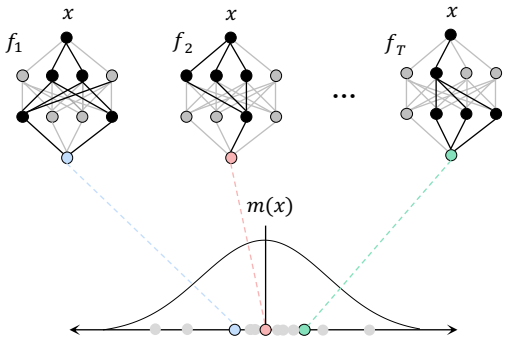


Figure 5.30: MC Dropout keeps dropout layers active during inference, resulting in different model estimations for each forward pass.

5.8.2 Model Training and Evaluation

In order to leverage MC Dropout in the context of NSE, the model described in Section 5.3 was extended by adding a dropout layer after each but the last hidden

layer. The dropout layers were kept active during testing. The final prediction $\hat{y}$ is calculated as the mean of all forward passes:

$$\hat{y} = \frac{1}{G} \sum_{i=1}^{G} y_i \tag{5.15}$$

where $G = 500$ is the number of forward passes and y_i is the prediction of the i-th forward pass. The standard deviation of the predictions is calculated as:

$$\sigma = \sqrt{\frac{1}{G} \sum_{i=1}^{G} (y_i - \hat{y})^2} \tag{5.16}$$

This standard deviation σ might provide a measure of the model's uncertainty in its predictions.

Figure 5.31 visualizes the prediction with the highest daily average deviation, the corresponding absolute voltage deviation $|\Delta_v|$, and the standard deviation σ_v of the model predictions for voltage magnitude. The general shape of σ_v and $|\Delta_v|$ visually correlate. If higher σ_v values indicate a higher estimation error $|\Delta_v|$, the models σ_v is a valid uncertainty estimator.

To formally assess the ability of the predicition's standard deviation to indicate model uncertainty, a statistical correlation between $|\Delta_v|$ and σ_v needs to be shown. An overall Pearson correlation of about 0.47 was found between the $|\Delta_v|$ and σ_v, indicating a moderate positive correlation. Furthermore the $|\Delta_v|$ for different percentiles of σ_v was calculated. Let $\mathcal{P}_{x,p}$ be the p-th percentile of x. The set of all σ_v values above the p-th percentile, is defined as

$$S_p = \{\sigma_{v,i} | \sigma_{v,i} > P_{\sigma_v,p}\} \tag{5.17}$$

The corresponding set of voltage magnitude deviation can be defined as

$$V_p = \{\Delta_{v,i} | \sigma_{v,i} \in S_p\} \tag{5.18}$$

Figure 5.32 plots the voltage mean absolute deviation $\bar{V_p}$ against the respective standard deviations, where

$$\bar{V_p} = \frac{1}{|V_p|} \sum_{v \in V_p} |v| \tag{5.19}$$

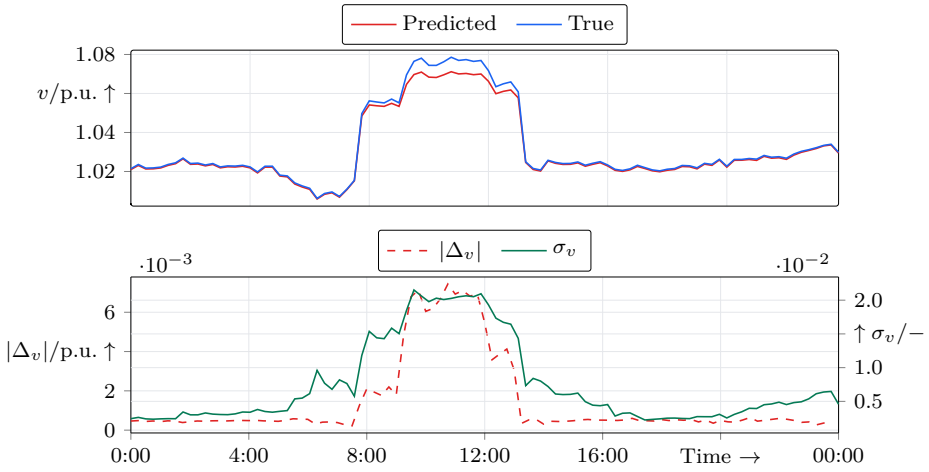


Figure 5.31: Prediction with the maximum daily average deviation (top), as well as absolute deviation, and standard deviation of prediction (bottom).

The plot illustrates, that the mean voltage deviation increases for higher standard deviation values, further underlining the uncertainty estimation capabilities of the model's standard deviation. Furthermore, the rate of change of the relation increases. Thus, the standard deviation is an even better predictor of model uncertainty for higher σ_v values.

Measured Evaluation

The synthetic application and evaluation have shown the potential of the dropout layers to facilitate uncertainty estimation. To analyze transferability of the approach the technique is applied using the measurement data.

The same round-robin procedure outlined in Section 5.4.3 was used to establish predictions for all measured buses. Figure 5.33 shows the worst-performing estimation across all nodes measured by the daily average deviation. A very similar profile can be seen when comparing the estimation deviation and the corresponding standard deviation of the model estimations. This indicates that the standard deviation is a valid indicator of the model error.

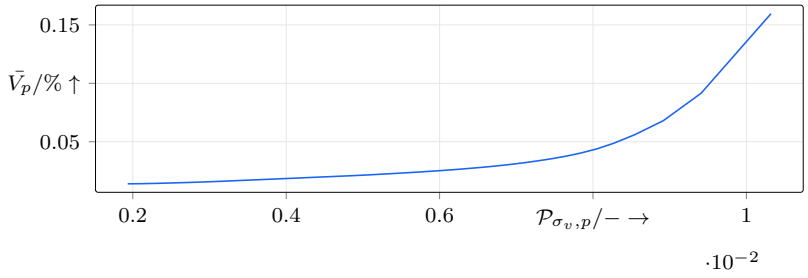


Figure 5.32: Mean voltage deviation for the set of predictions on synthetic data with a greater standard deviation than $P_{\sigma_v,p}$.

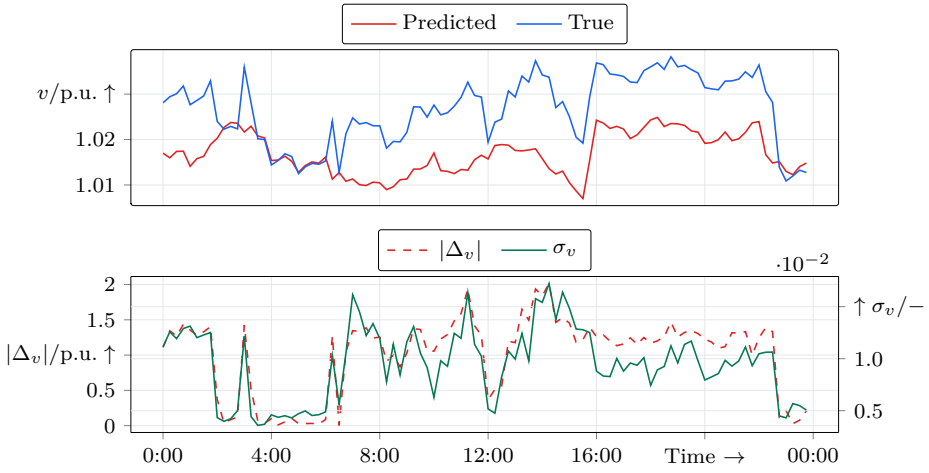


Figure 5.33: Prediction of the model with dropout layer (top) as well as the prediction error and the corresponding prediction standard deviation (bottom).

The overall Pearson correlation is about 0.65, indicating a moderate to strong positive correlation and an even higher overall correlation as for the synthetic data set. Similarly as in Figure 5.32 for the synthetic data, the mean average deviation $\bar{V}_p$ for the different standard deviation percentiles is depicted in Figure 5.34. The graph clearly shwos the correlation between the estimation standard deviation and the average voltage deviation. Given the results it can be concluded that the methodology is transferable to the measurement data.

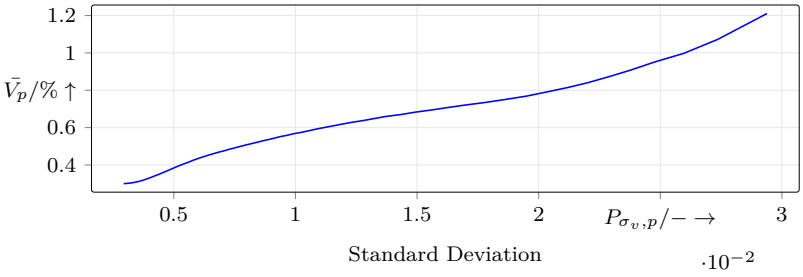


Figure 5.34: Mean voltage deviation for the set of predictions on measured data with a greater standard deviation than $P_{\sigma_v,p}$.

A subsequent analysis investigates wether the standard deviation is bus-specific or input sample specific. In other words: Is the standard deviation high for each bus simultaneously due to specifics of the input sample, or high for individual buses and therefore localizable? Thus, the Pearson correlation of the standard deviation between all bus pairs is calculated to investigate whether high standard deviation values correlate across nodes. For two buses i and j, let $\sigma_{v,i}(t)$ and $\sigma_{v,j}(t)$ represent their respective standard deviations at time step t. The Pearson correlation coefficient $\rho_{i,j}$ between their standard deviations is calculated as:

$$\rho_{i,j} = \frac{\sum_{t=1}^{T}(\sigma_{v,i}(t) - \bar{\sigma}_{v,i})(\sigma_{v,j}(t) - \bar{\sigma}_{v,j})}{\sqrt{\sum_{t=1}^{T}(\sigma_{v,i}(t) - \bar{\sigma}_{v,i})^2}\sqrt{\sum_{t=1}^{T}(\sigma_{v,j}(t) - \bar{\sigma}_{v,j})^2}} \tag{5.20}$$

where $\bar{\sigma}_{v,i}$ and $\bar{\sigma}_{v,j}$ are the mean standard deviations for buses i and j respectively across all time steps T. The correlation matrix shown in Figure 5.35 visualizes these coefficients for all bus pairs.

	B01	B32	B18	B11	B38	B13	B34	B44	B23	B31	B29	B30	B21	B16	B14	B35	B22	B07	B43	B42
B01	1.00	0.87	0.88	0.87	0.88	0.88	0.87	0.88	0.88	0.87	0.87	0.86	0.86	0.87	0.86	0.87	0.87	0.86	0.87	0.83
B32	0.87	1.00	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.98	0.98	0.97	0.97	0.97
B18	0.88	0.98	1.00	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.97	0.97	0.98	0.97	0.98	0.96
B11	0.87	0.98	0.98	1.00	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97
B38	0.88	0.97	0.98	0.98	1.00	0.98	0.97	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.98	0.96
B13	0.88	0.97	0.97	0.98	0.98	1.00	0.97	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97
B34	0.87	0.97	0.97	0.98	0.97	0.97	1.00	0.97	0.97	0.98	0.97	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96
B44	0.88	0.98	0.98	0.98	0.98	0.98	0.97	1.00	0.98	0.98	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.98	0.96
B23	0.88	0.98	0.98	0.98	0.98	0.98	0.97	0.98	1.00	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.97
B31	0.87	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	1.00	0.97	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97
B29	0.87	0.97	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.97	1.00	0.98	0.98	0.98	0.97	0.97	0.98	0.97	0.98	0.97
B30	0.86	0.97	0.97	0.97	0.97	0.97	0.96	0.97	0.98	0.97	0.98	1.00	0.97	0.98	0.97	0.97	0.98	0.97	0.97	0.97
B21	0.86	0.97	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.97	1.00	0.97	0.97	0.97	0.98	0.97	0.98	0.97
B16	0.87	0.97	0.97	0.98	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.97	1.00	0.97	0.98	0.98	0.97	0.98	0.97
B14	0.86	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	1.00	0.97	0.97	0.97	0.98	0.97
B35	0.87	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.98	0.97	1.00	0.97	0.97	0.97	0.97
B22	0.87	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97	1.00	0.97	0.98	0.97
B07	0.86	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	1.00	0.98	0.97
B43	0.87	0.97	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.97	0.98	0.98	1.00	0.97
B42	0.83	0.97	0.96	0.97	0.96	0.97	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	1.00

Figure 5.35: Correlation matrix of standard deviations across different buses, showing how uncertainty estimates are related between buses.

Given that the lowest overall correlation value is 0.83, it can be concluded that standard deviation values are highly correlated. This shows that the standard deviation is not bus-specific but input sample-specific. A high standard deviation, in turn, can be used to classify uncertainty for all estimations of a given input sample and can't be localized to bus specific errors.

5.8.3 Probabilistic State Estimation

The previous analysis has demonstrated that MC Dropout provides valuable uncertainty information for unmeasured bus voltage estimations. Building on this insight, this section extends the approach from uncertainty detection to a probabilistic state estimation framework. Instead of producing single-point predictions with separate uncertainty indicators, a method is proposed that generates probability distributions of possible voltage values.

The main approach is to associate the estimation standard deviation with a historical error distribution. Overlaying the error distribution onto the model estimation results in a probabilistic estimation. To generate the error distributions, we first analyze the distribution of standard deviations across all buses and time steps. For each time step t, the mean standard deviation across all buses is calculated:

$$\bar{\sigma}_v(t) = \frac{1}{N} \sum_{i=1}^{N} \sigma_{v,i}(t) \tag{5.21}$$

where N is the number of buses. These mean standard deviations are then grouped into bins using a histogram approach:

$$B_k = \{\bar{\sigma}_v(t) \mid b_k \leq \bar{\sigma}_v(t) < b_{k+1}\} \tag{5.22}$$

where B_k represents the k-th bin and b_k, b_{k+1} are its boundaries. The resulting histogram is depicted in Figure 5.36, where F represents the frequency of the bin. The size of the bars depicts the number of input samples whose mean standard deviation falls within the bin's boundaries.

For each bin B_k, all corresponding estimation errors $\Delta_{v,i}(t)$ where $\bar{\sigma}_v(t) \in B_k$ are collected. Bin-specific error distributions can thus be constructed that characterize the relationship between the model's uncertainty (represented by σ_v) and its actual estimation errors. The probability density histogram of the error distributions are shown in Figure 5.37. The normalized frequencies of the histograms are denoted with

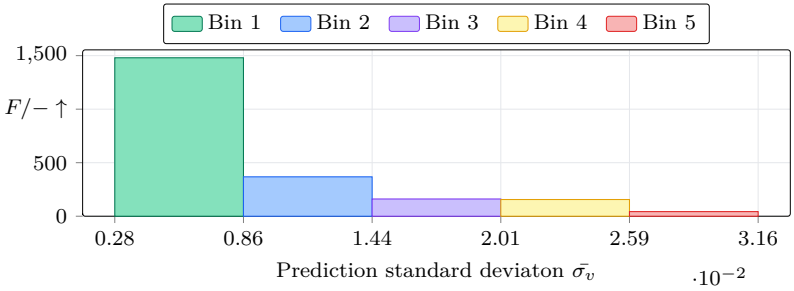


Figure 5.36: Histogramm of the model's standard deviation.

$\tilde{F}$. The distributions are tilted to the right, showing a tendency to underestimate the voltage magnitude. As expected, higher deviations can be seen in the higher bins, representing higher model uncertainty.

Thus, each estimation can be associated to a bin and the bin to a historical error distribution. The error distribution can be used to estimate prediction ranges for the model estimation. Three confidence intervals for the individual error distributions were calculated. The confidence intervals used are based on one, two and three standard deviations. This results in percentile ranges from P_{16} to P_{84}, $P_{2.5}$ to $P_{97.5}$ and $P_{0.5}$ to $P_{99.5}$. For each error distribution the confidence interval boundaries have been calculated and are noted in Table 5.7.

Table 5.7: Confidence intervals for error distributions by standard deviation bin

	CI 68% (P_{16} – P_{84}) $\lvert\Delta_v\rvert$ in %	**CI 95%** ($P_{2.5}$ – $P_{97.5}$) $\lvert\Delta_v\rvert$ in %	**CI 99%** ($P_{0.5}$ – $P_{99.5}$) $\lvert\Delta_v\rvert$ in %
Bin 1	(-0.12 - 0.26)	(-0.39 - 0.47)	(-0.55 - 0.66)
Bin 2	(-0.34 - 0.56)	(-0.81 - 1.02)	(-0.96 - 1.41)
Bin 3	(-0.59 - 0.77)	(-1.11 - 1.25)	(-1.26 - 1.67)
Bin 4	(-0.72 - 1.10)	(-1.32 - 1.60)	(-1.53 - 1.80)
Bin 5	(-0.76 - 1.50)	(-1.53 - 2.00)	(-1.69 - 2.20)

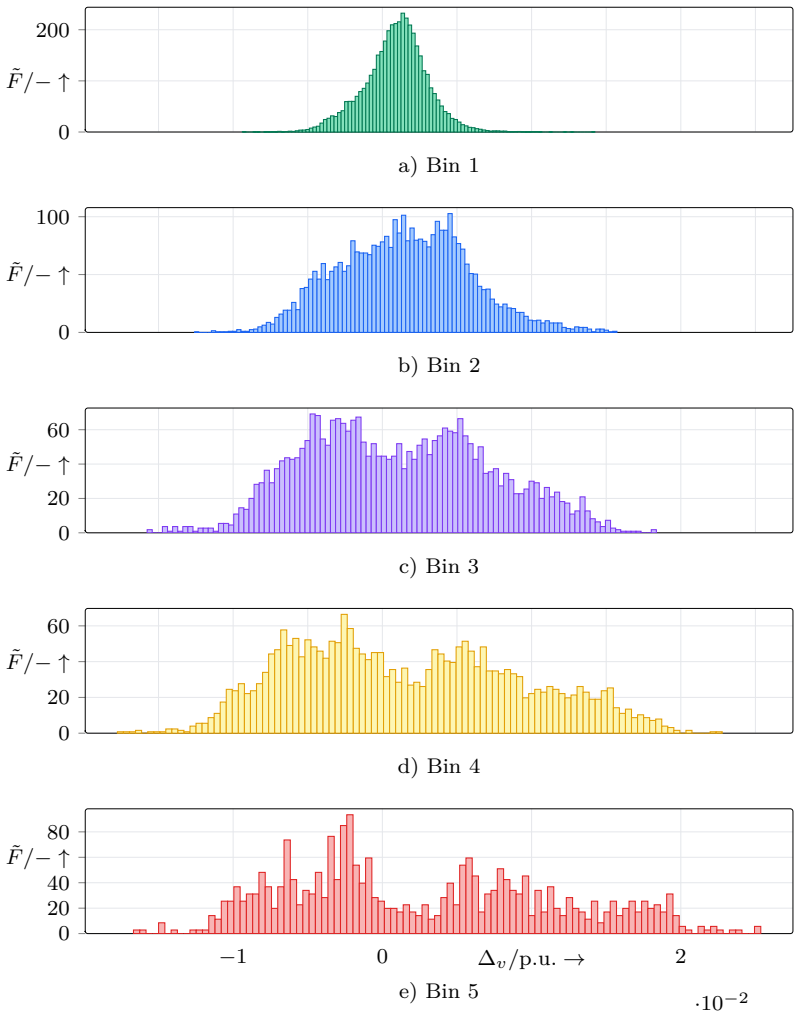


Figure 5.37: Probability density histogram of estimation deviation grouped by the model's standard deviation.

Given the confidence intervals derived from the error distributions, a complete procedure for probabilistic state estimation can be formulated. The process consists of the following steps:

1. **Model Execution:** Run the neural network model with active dropout layers to obtain both the voltage estimation $\hat{v}$ and the corresponding standard deviation σ_v through multiple forward passes.
2. **Uncertainty Classification:** Calculate the mean standard deviation $\bar{\sigma}_v$ across all buses for the current input sample and classify it into one of the five predefined bins (as shown in Figure 5.36).
3. **Confidence Interval Selection:** Based on the assigned bin, retrieve the corresponding error distribution parameters from Table 5.7.
4. **Prediction Range Calculation:** Apply the confidence interval boundaries to the point estimation to generate the probabilistic prediction ranges:

$$v_{lower} = \hat{v} + P_{lower} \tag{5.23}$$

$$v_{upper} = \hat{v} + P_{upper} \tag{5.24}$$

   where P_{lower} and P_{upper} are the lower and upper percentile values from the selected confidence interval.

This procedure transforms point predictions into probabilistic estimations with uncertainty bounds. An exemplary probabilistic prediction is shown in Figure 5.38. The colors indicate the result of the uncertainty classification of the estimation, while the opacity represent the different confidence intervals. Compared to a simple prediction line plot, the stochastic prediction visualization transports much more information about the estimation ranges and historical associated probabilities.

The calculated error ranges are customized for the specific model and based on historical measurements. Consequently, if the model changes, the error distribution and resulting confidence intervals must be recalculated. Furthermore, if the underlying load and consumption patterns change the error distribution is expected to change, which should also trigger a reparameterization of the error distribution. Since changing the model is a known event, a subsequent reparameterization is straightforward. Knowing when the underlying load and consumption patterns have changed significantly enough to necessitate an error distribution update is not. Further research is necessary. A possible approach is to monitor the model uncertainty over time and use the information to trigger a reparameterization.

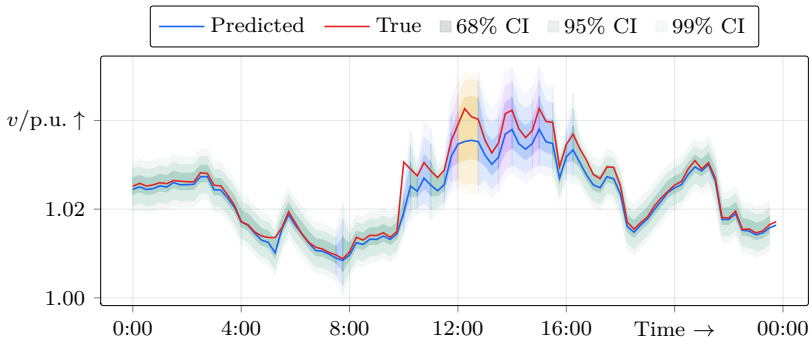


Figure 5.38: Probabilistic prediction with uncertainty ranges with colors indicating the different error distributions, based on the model uncertainty.

In summary, the proposed methodology gives system operators a more comprehensive understanding of the possible network states. The uncertainty ranges could be used to implement a traffic light system, indicating the likelihood of exceeding voltage limits. Different severity levels (e.g. green, yellow, and red) could be defined given the percentage chance of the voltage reaching levels outside of defined voltage limits. Furthermore, the probabilistic voltage information can inform a set of different operational measures based on the severity of the limit violations and the respective likelihood.

5.9 Summary

This chapter investigates using neural networks for distribution system state estimation, referred to as NSE. The NSE approach is benchmarked against a conventional WLS state estimator. The ANN model is trained on the synthetic data set of the real distribution grid discussed in Chapter 3.

Initially, the performance of both models is evaluated on the synthetic, clean dataset and the synthetic dataset with added noise. While the ANN model performs slightly worse on the clean synthetic dataset variant, it handles the introduction of noise significantly better than the WLS model. Regarding voltage angle and line utilization estimation the ANN model performs very well while the WLS model struggles.

To investigate the transferability of the synthetically trained model to the field, its performance using actual measurement data is evaluated. Round-robin training and evaluation is proposed to assess the performance of the NSE model given the limited set of real measurements. Due to the limited availability of measurement data, only voltage magnitude estimations can be evaluated. Overall, both models perform well on the measurement data set. While showing adequate performance overall, noticeable performance degradation can be seen by comparing the deviation levels to the synthetic evaluation scenario. This shows that while the synthetically trained ANN model is transferable, the performance on synthetic data cannot be used to gauge performance in a field environment.

Computational speed comparisons reveal that the NSE model's inference is orders of magnitude faster than the WLS algorithm's execution. Analysis of the NSE model's performance across different buses highlights the variability of estimation performance across different training runs. This is a challenge concerning reproducibility and the interpretation of model estimation performance.

The chapter also examines the NSE model's robustness to changes in grid utilization, finding that the model's accuracy degrades when encountering load or generation levels outside the range seen during training. Fine-tuning the model on a mix of regular and extreme data samples is proposed as a way to widen its applicability while retaining performance on the original data distribution.

Finally, the use of MC Dropout is explored to quantify the NSE model's uncertainty. The model's prediction variance, obtained through multiple forward passes with dropout, correlates well with its estimation error, providing uncertainty intervals alongside point predictions. Building on this approach, a way of using the dropout layers for probabilistic state estimation is presented. Instead of simple point predictions, the probabilistic state estimation delivers an estimation containing different estimation ranges based on the prediction variance and historical errors.

6 System Overview

During Chapter 3 to Chapter 5 a multitude of subsystems have been discussed and developed that make up an exemplary NSE implementation. The subsystems include synthetic data generation, measurement processing and, the stochastic NSE. This chapter gives a summarizing overview of the role of each of the subsystems and their integration into a larger system. Furthermore, some of the practical implementation considerations are discussed.

6.1 System Components

As part of this thesis, multiple functional components have been developed and analyzed to address the challenges of monitoring modern distribution grids. Practical considerations like the specific API design of the components, the integration of these components into a SCADA system, and field deployment are out of the scope of this thesis. However, this section aims to give a summarizing overview of the individual components implemented in this thesis, illustrating how they integrate into a combined system. The components can be grouped into two main parts. The first is the overall model training pipeline, depicted in Figure 6.1, which enables the creation of supervised learning models. The second consists of the operational DSSE modules, shown in Figure 6.2, which utilize these models for real-time grid monitoring.

Foundational to the model training pipeline is the accurate representation of the physical grid model. As the degree of digitization increases, more information about the grid is readily available. However, the focus currently lies mostly on the physical grid infrastructure and related assets. For simulating grid usage, information on connected load and generation systems is required to form an integrated digital model of the current grid. The process of modeling a synthetic grid and simulating grid usage is discussed in Chapter 3. Building on top of this integrated grid model, a scenario generator can provide different simulation scenarios. A status quo scenario relies on the current grid model and aims to simulate grid usage and by extension grid states that represent typical operation conditions of the physical system. Additional fine-tuning scenarios can be generated to increase the robustness of the model. Within these scenarios, expected future changes in consumption and generation patterns can be modeled. Each scenario forms the basis for a simulation run, which generates a labeled dataset representing the scenario-specific circumstances.

Agent-based power system simulators like SIMONA are best suited as the simulation environment. These simulators are able to simulate not only the behavior of a variety of different assets but also the interaction between them. This gets increasingly important as the penetration of energy management systems in modern distribution grids increases. The resulting labeled data sets form the basis for ML model training. As part of the training pipeline, a defined preprocessing stage prepares the data for subsequent model training. To make the trained models accessible, they can be placed and versioned in a central model hub. Versioning the models would allow for a rollback if unexpected behavior is observed and the possibility of tracking changes in model performance over time.

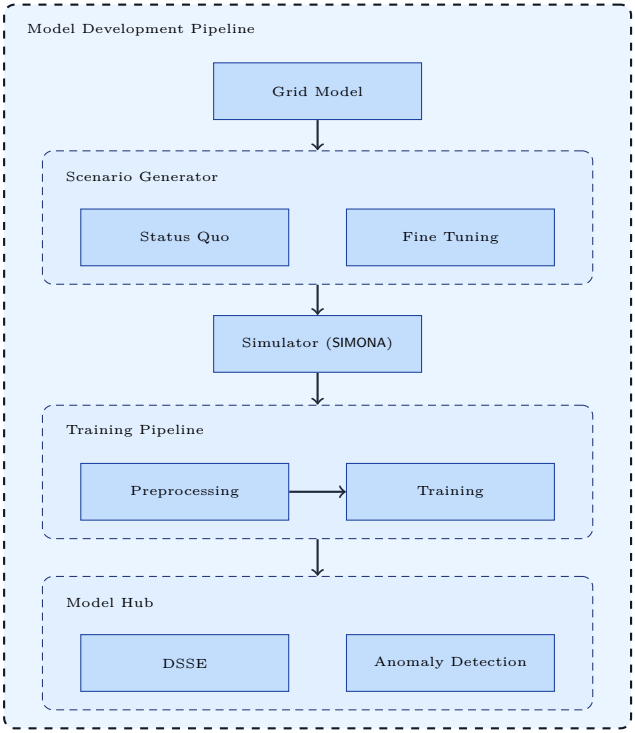


Figure 6.1: Component overview of the model development pipeline.

From an operational perspective, the DSSE process, depicted in Figure 6.2, starts with raw, real-time field measurements recorded and transferred to the SCADA system. To address inherent data quality issues present in field measurements, a measurement preprocessing phase is employed. The measurement preprocessing phase comprises the Anomaly Detection and Signal Reconstruction processes discussed in Section 4.3 and Section 4.4. Both processes would utilize the respective models from the Model Hub discussed earlier. The anomaly detection flags missing or anomalous measurements. To enhance data quality, the signal reconstruction process then utilizes healthy measurements to reconstruct the anomalous measurements. The preprocessed input data forms the basis of the actual SE process. Using a probabilistic state estimation method, discussed in Section 5.8, the current grid state including estimation ranges is calculated. Given this calculated state, the grid's health can be monitored and visualized in the SCADA system, supporting operational use cases like congestion management. If available, short-term operational measures can be taken to alleviate potential grid congestions.

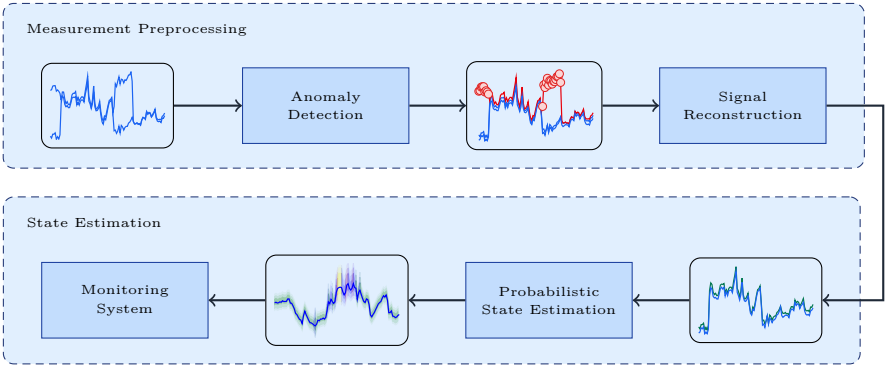


Figure 6.2: Component overview of the DSSE process.

6.2 Practical Considerations

This section discusses some of the practical considerations for the implementation of the NSE model. While not exhaustive it touches on some of the important challenges on setting up and operating the NSE.

Digital Infrastructure

The state of digitization and the lack of experience of most distribution grid operators with data-driven methods pose a challenge for their implementation. Considerable amounts of effort have to be expended for setting up the general infrastructure and training staff on ML techniques as well as software development to implement the functionality from the ground up. Given the considerable effort and the comparable performance of classical methods the supervised ML methods might not seem like the most effective choice. Instead, the implementation of data-driven methods should be viewed as part of a more foundational shift in grid operation paradigms. This shift should move towards increased digitization, particularly through the implementation of digital twin-like systems. Such systems would serve as a central hub, providing a continuously updated and readily accessible source for all grid-related information. The development of standardized APIs for these systems would enable the creation of standardized data-driven solutions. However, standardization efforts should not be limited to the overall data source. They should also encompass the interfaces of all functional components discussed in the previous section. This ensures that each component remains modular and interchangeable, facilitating future upgrades and improvements to the system. Furthermore, specialized software can be built by software-focused companies alleviating the need for grid operators for extended in-house software development. If a suitable data platform exists, each application case becomes merely a plugin of a broader system.

Automated Retraining

As the training of a model represents a set of operation conditions and a network configuration that might change over time, effectively retraining or subsequently fine-tuning a model is an important consideration. To ease the operation of such a system a high degree of automation is desirable and achievable. Specific changes within a central database, like the installation of a new line, could automatically trigger a model training workflow. After retraining the model a streamlined way of

testing and deploying the model is needed. To ensure operational security of tests need to be developed and employed that test the models under all operating conditions. Only after succesfully passing the tests, the model can then be operated.

Handling Different Time Resolutions

One of the challenges mentioned in Section 2.4 is the diversity of time resolutions due to differences in the deployed measurement infrastructure. One possible option for dealing with that is to ignore the missing measurements of the devices with a coarser resolution and reconstruct them with the signal reconstruction model. A second option would be to train multiple DSSE models with different input configurations, corresponding to the availability of measurements.

Topology Changes

The DSSE model is specific to the grid topology it was trained on. When the topology changes, the model is likely not viable anymore. A simple solution is to initiate a retraining process. For planned grid extensions this solution is feasible. If the topology changes frequently due to switching operations different models corresponding to the different topologies can be stored in a central repository. The model to be used can then be selected based on the current topology.

7 Summary and Outlook

By addressing the research questions that were posed in Chapter 1, the results of this thesis are summarized in Section 7.1. Building upon this summary, Section 7.2 then outlines promising and specific directions for future research.

7.1 Summary

The integration of renewable energies and sector coupling advances necessitate advancements in active grid monitoring and operational control in DGs. All operational measures build upon a functional DSSE to establish grid observability. Among DSSE approaches, supervised learning methods using ANN, termed NSE, have gained traction. These methods learn a mapping from available measurements to the complete grid state. The fundamental challenge lies in training these supervised models: they require complete grid state data (including the unmeasured parts) for learning, precisely what is unavailable in real-world operation. Current research predominantly circumvents this by training and validating NSE models using purely synthetic data generated from standardized test grids and simulated load/generation profiles. While useful for algorithmic development, this raises crucial questions about the practical applicability of such models in real distribution grids. This thesis is situated within this context, exploring the potential of leveraging digital representations of real grids. It investigates whether an NSE model trained on data synthesized from a digital representation of a specific, real grid is transferable to field application. Furthermore, practical hurdles concerning data quality issues present in field measurements are investigated. Finally, acknowledging the inherent imprecision in estimating unmeasured states and the lack of transparency in ANN models, the thesis addresses the need for uncertainty quantification and proposes an ANN based stochastic DSSE.

Research Question 1: *Is a NSE model trained on a synthetic representation of a real grid transferable to field operation given our limited information about real distribution grids?*

To investigate a synthetic grid model of a real MV distribution grid is developed. Based on this synthetic grid model, the grid behavior is simulated using an energy system simulator (see Chapter 3). In Chapter 5, the resulting synthetic data set is used to train an ANN and benchmarked against a conventional WLS model.

To evaluate transferability the estimation performance of the ANN using real field measurements is analyzed. Due to the limited availability of measurement data, only voltage magnitude estimations can be evaluated. The ANN shows adequate performance on the measurement data set establishing a general transferability of the model. Importantly, the performance on the measurement data is significantly worse than on the synthetic data set. This performance degradation shows that model performance on synthetic data is not necessarily indicative of performance during field application. The ANN struggles to outperform the WLS model although overall performance is similar. This investigation shows general transferability to the current state grid usage. To increase the model's robustness to changes in grid usage fine-tuning has shown promising results. By adding samples that widen the training data distribution, the model shows better estimation results in a wider data range.

Research Question 2: *Can synthetically trained ANNs be leveraged for anomaly detection and signal reconstruction of field measurements in the context of NSE?*

Unlike synthetic signals, real-world measurement data suffers from quality issues like corruption and missing values. These problems degrade estimation performance and challenge fixed-input models like ANN and are addressed in Chapter 4. To detect anomalies, an DLAE model is developed. The approach leverages the autoencoder's ability to reconstruct valid data with small errors while producing larger reconstruction errors for anomalous data. Upon exceeding a reconstruction error threshold a signal is considered anomalous. The model was trained on the synthetic data set and showed very good performance on the synthetic data set. Evaluating the model on the field measurements proves more challenging since the measurement data is unlabeled so anomalies can only be assumed. Upon qualitative assessment of the detection mechanism on assumed anomalies, the model proves successful.

Anomalous or missing data in the context of grid monitoring can be problematic in downstream applications. Based on the anomaly detection model a signal reconstruction method for field measurements is analyzed. The method first generates pseudo-measurements by weighted averaging valid neighbors identified via graph traversal. An autoencoder then iteratively refines these values. The method proved succesful in reconstructing identified anomalies within the field measurement set.

Research Question 3: *How can the estimation uncertainty of NSE models be quantified and used for stochastic SE?*

Limited transparency, the difficulty in understanding ML model decision-making, impedes their use in critical infrastructure. This is a well-founded concern for applications like DSSE, which informs operational actions and monitors electrical equipment status. Section 5.8 addresses the associated research question. Uncertainty estimation of the ANN is demonstrated with the usage of MC Dropout. The variance of multiple dropout passes is shown to quantify the estimation error and therefore estimation uncertainty. The MC Droput approach is extended to a stochastic SE. Instead of outputting only point predictions, this method uses the dropout variance and historical error data to provide estimates with associated uncertainty ranges.

7.2 Outlook

Grid Models and Generalization: The developed models and their performance characteristics are assessed based on the specific MV grid that was available. To validate and generalize the results different grid models with different topologies and different load and consumer compositions should be investigated. Especially potential differences in performance characteristics between smaller and larger grids should be investigated. Furthermore, the effect of meshed grids should be analyzed.

Model Architectures and Physics Integration: With respect to model architectures, MLPs have been used for the DSSE and the encoder and decoder modules of the autoencoder. Although some other model architectures like long short-term memory and graph neural networks have not improved upon the MLP performance for this data set, more research should be done to confirm these results. Since the DSSE model is not bound by power flow equations, it can output physically infeasible states. This can be improved by integrating the power flow equations into the cost function during training.

Measurement Data Limitations and Validation: The evaluation of model performance on real-world data was constrained by the available measurements, notably the lack of synchronized voltage angle and direct line current measurements. Although the ANN demonstrated superior performance compared to WLS for estimating these quantities on synthetic data, these improvements could not be empirically validated. Future research necessitates access to richer datasets incorporating detailed, synchronized measurements (e.g., from μPMU or line sensors). This

will allow for the validation of whether the synthetically trained models maintain their superior performance for angle and line utilization estimation when applied to real measurement data. Field tests should be designed to capture diverse operating conditions, including those where accurate angle and current data are vital (e.g., near congestion limits, during high renewables fluctuations).

Fine-Tuning Validation and Automation: It was shown that fine-tuning could improve the model's robustness to changes in the distribution of the synthetic data. However, these results should be confirmed using real-world field measurements. For this investigation, datasets that encompass significant, quantifiable shifts in load or generation patterns are necessary. Examples include the commissioning of large industrial loads, significant integration of renewable energy resources, or widespread adoption of new technologies like electric vehicles or heat pumps. As more measurement infrastructure is installed in the distribution grid and the adoption speed increases, more data will be available to initiate the specific research. Furthermore, practical implementation requires investigating the optimal frequency for fine-tuning and developing methods for detecting significant changes in the underlying statistical properties of the measurement data that would trigger model retraining or fine-tuning processes.

Anomaly Detection Extension: While anomaly detection and signal reconstruction have worked well for voltage magnitude it might not work as effectively for detecting anomalous power values. As there is much less correlation between power values as opposed to voltage magnitudes of surrounding nodes, anomalies can be harder to detect. The effectiveness of the anomaly detection and signal reconstruction for other signals should be analyzed.

Adapting to Grid Topology Changes: A significant limitation of the currently trained NSE models is their specificity to the grid topology they were trained on. Performance is expected to degrade substantially if the topology changes due to events like switching operations, line maintenance, or faults. Addressing topology changes requires dedicated research. Promising research areas include retraining upon detected topology changes, pretraining different models for likely switch positions, including switch positions in the NSE model, or leveraging the inherent topology representation of graph neural networks.

Addressing Performance Variability: A notable challenge identified is the variability in NSE performance across different training runs. Models achieving similar performance metrics on the synthetic validation data can demonstrate significantly different accuracy when applied to real measurement data. Most likely, slight dif-

ferences in the local minima found by the optimizer generalize differently to the measurement data. Future research should investigate methods to promote convergence towards minima that generalize better. Different regularization methods, for example including a loss term penalizing power flow inconsistencies, could alleviate the issue. Furthermore, improved validation strategies that better predict real-world generalization during the training phase might help decrease model performance variance.

Bibliography

[1] Bundesverband der Energie- und Wasserwirtschaft e.V. (BDEW), "Die Energieversorgung 2024," Berlin, Tech. Rep., Dec. 2024. [Online]. Available: https : / / www . bdew . de / media / documents / 2024 _ 12 _ 18 _ Die _ Energieversorgung_2024_Final.pdf (visited on 04/07/2025).

[2] M. S. Alam and S. A. Arefifar, "Energy Management in Power Distribution Systems: Review, Classification, Limitations and Challenges," *IEEE Access*, vol. 7, pp. 92 979–93 001, 2019, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2927303.

[3] S. R. Salkuti, "Challenges, issues and opportunities for the development of smart grid," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 2, p. 1179, Apr. 2020, ISSN: 2088-8708, 2088-8708. DOI: 10.11591/ijece.v10i2.pp1179-1186.

[4] *Gesetz über die Elektrizitäts- und Gasversorgung (Energiewirtschaftsgesetz - EnWG)*, German Federal Law, ğ 14a Netzorientierte Steuerung von steuerbaren Verbrauchseinrichtungen und steuerbaren Netzanschlüssen; Festlegungskompetenzen.

[5] VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V., "Hochautomatisierung von Nieder- und Mittelspannungsnetzen," Offenbach am Main, VDE Studie, Jul. 2023, Available online or from VDE. [Online]. Available: https : / / www . vde . com / de / etg / arbeitsgebiete / v2 / hochautomatisierung - von - nieder – und - mittelspannungsnetzen (visited on 02/05/2025).

[6] S. Vitiello, N. Andreadou, M. Ardelean, *et al.*, "Smart Metering Roll-Out in Europe: Where Do We Stand? Cost Benefit Analyses in the Clean Energy Package and Research Trends in the Green Deal," *Energies*, vol. 15, no. 7, p. 2340, Mar. 2022, ISSN: 1996-1073. DOI: 10.3390/en15072340.

[7] A. Abur and A. G. Expósito, *Power System State Estimation*. CRC Press, Mar. 2004, ISBN: 978-0-203-91367-3. DOI: 10.1201/9780203913673.

[8] W. Sulaimon and I. Habiballah, "Recent Advances in Distribution System State Estimation Algorithms: From Model-based to Data-driven Approach," Engineering, Preprint, Apr. 2023. DOI: 10.20944/preprints202304.1203.v1.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (Adaptive Computation and Machine Learning). Cambridge, Massachusetts: The MIT Press, 2016, ISBN: 978-0-262-03561-3.

[10] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-Driven Learning-Based Optimization for Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4805, Nov. 2019, ISSN: 0885-8950, 1558-0679. DOI: 10.1109/TPWRS.2019.2909150.

[11] O. Kundacina, M. Cosovic, and D. Vukobratovic, "Robust and Fast Data-Driven Power System State Estimator Using Graph Neural Networks," 2022. DOI: 10.48550/ARXIV.2206.02731.

[12] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Graph Neural Network-Based Distribution System State Estimators," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 12, pp. 11 630–11 639, Dec. 2023, ISSN: 1551-3203, 1941-0050. DOI: 10.1109/TII.2023.3248082.

[13] J. Ostrometzky, K. berestizshevsky, A. Bernstein, *et al.*, "Physics-Informed Deep Neural Network Method for Limited Observability State Estimation," 2019. DOI: 10.48550/ARXIV.1910.06401.

[14] V. ETG, *The Digital Twin in the Network and Electricity Industry*, 2023.

[15] J. Xu, Y. Jin, T. Zheng, *et al.*, "On State Estimation Modeling of Smart Distribution Networks: A Technical Review," *Energies*, vol. 16, no. 4, p. 1891, Feb. 2023, ISSN: 1996-1073. DOI: 10.3390/en16041891.

[16] A. Pillay, S. Prabhakar Karthikeyan, and D. P. Kothari, "Congestion management in power systems A review," *International Journal of Electrical Power & Energy Systems*, vol. 70, pp. 83–90, Sep. 2015, ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2015.01.022.

[17] *DIN EN 50160, Voltage characteristics of electricity supplied by public electricity networks*, 2020. DOI: 10.31030/3187943.

[18] D. Croteau and O. Carre, "Distribution state estimation: Outcomes from a field implementation aimed at tackling MV mastering in the presence of distributed energy resources (DER)," *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 1715–1717, Oct. 2017, ISSN: 2515-0855. DOI: 10.1049/oap-cired.2017.0420.

[19] M. Viania Sebastian, M. Caujolle, B. Goncer Maraver, *et al.*, "LV state estimation and TSODSO cooperation tools: Results of the French field tests in the evolvDSO project," *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 1883–1887, Oct. 2017, ISSN: 2515-0855. DOI: 10.1049/oapcired.2017.0410.

[20] L. Jendernalik, T. Wiedemann, P. Noglik, *et al.*, "The German demo inside GRID4EU: Field experience, lessons learnt and some aspects about cost-benefit of innovative solutions," in *CIRED Workshop 2016*, Helsinki, Finland: Institution of Engineering and Technology, 2016, 70 (4 .)–70 (4 .) ISBN: 978-1-78561-202-2. DOI: 10.1049/cp.2016.0670.

[21] M. Ramezani, M. Etemadizadeh, and H. Falaghi, "Reactive power management in distribution networks in the presence of distributed generation sources based on information gap decision theory," *Sustainable Energy, Grids and Networks*, vol. 39, p. 101 470, Sep. 2024, ISSN: 23524677. DOI: 10.1016/j.segan.2024.101470.

[22] A. Schurtz, *Verfahren zur Zustandsschätzung und ihr Beitrag zum Engpassmanagement in Mittelspannungsnetzen* (Dortmunder Beiträge zu Energiesystemen, Energieeffizienz und Energiewirtschaft). Düren: Shaker Verlag, 2020, vol. 19, ISBN: 9783844076387.

[23] D. Tomaselli, D. Most, E. Sinani, *et al.*, "Leveraging Prosumer Flexibility to Mitigate Grid Congestion in Future Power Distribution Grids," *Energies*, vol. 17, no. 17, p. 4217, Aug. 2024, ISSN: 1996-1073. DOI: 10.3390/en17174217.

[24] A. Vanselow, S. Krahl, A. Moser, *et al.*, "Simulation and analysis of a congestion management utilizing load-side flexibilities within the distribution grid," in *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Grenoble, France: IEEE, Oct. 2023, pp. 1–5, ISBN: 9798350396782. DOI: 10.1109/ISGTEUROPE56780.2023.10407765.

[25] E. Handschin, *Elektrische Energieübertragungssysteme. Hauptbd.* 2. Aufl. Heidelberg: Hüthig, 1987, ISBN: 978-3-7785-1401-6.

[26] R. E. Larson, W. F. Tinney, and J. Peschon, "State Estimation in Power Systems Part I: Theory and Feasibility," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 3, pp. 345–352, Mar. 1970, ISSN: 0018-9510. DOI: 10.1109/TPAS.1970.292711.

[27] P. H. Nguyen, G. K. Venayagamoorthy, W. L. Kling, *et al.*, "Dynamic state estimation for distribution networks with renewable energy integration," *International Journal of Smart Grid and Clean Energy*, vol. 2, no. 3, pp. 307–315, 2013, ISSN: 23154462. DOI: 10.12720/sgce.2.3.307-315.

[28] J. Zhao, J. Qi, Z. Huang, *et al.*, "Power System Dynamic State Estimation: Motivations, Definitions, Methodologies, and Future Work," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3188–3198, Jul. 2019, ISSN: 0885-8950, 1558-0679. DOI: 10.1109/TPWRS.2019.2894769.

[29] Y. Lin and A. Abur, "Robust State Estimation Against Measurement and Network Parameter Errors," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4751–4759, Sep. 2018, ISSN: 1558-0679. DOI: 10.1109/TPWRS.2018.2794331.

[30] M. Celik and A. Abur, "A robust WLAV state estimator using transformations," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 106–113, Feb. 1992, ISSN: 1558-0679. DOI: 10.1109/59.141693.

[31] A. Gjelsvik, S. Aam, and L. Holten, "Hachtel's Augmented Matrix Method - A Rapid Method Improving Numerical Stability in Power System Static State Estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 11, pp. 2987–2993, Nov. 1985, ISSN: 0018-9510. DOI: 10.1109/TPAS.1985.318939.

[32] R. Brandalik, D. Henschel, and W. H. Wellssow, "A Computationally Efficient State Estimation Algorithm for the Supervision of Low Voltage Grids," in *2018 Power Systems Computation Conference (PSCC)*, Dublin: IEEE, Jun. 2018, pp. 1–7, ISBN: 978-1-910963-10-4. DOI: 10.23919/PSCC.2018.8442462.

[33] Y. Boukili, M. M. Ayiad, H. Moayyed, *et al.*, "Robust State Estimation Model for Low Voltage Distribution Networks in the Presence of Multiple Gross Errors," *IEEE Access*, vol. 11, pp. 42 403–42 415, 2023, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3270386.

[34] C. Hernandez and P. Maya-Ortiz, "Comparison between WLS and Kalman Filter method for power system static state estimation," in *2015 International Symposium on Smart Electric Distribution Systems and Technologies (EDST)*, Vienna, Austria: IEEE, Sep. 2015, pp. 47–52, ISBN: 978-1-4799-7736-9. DOI: 10.1109/SEDST.2015.7315181.

[35] A. M. Kettner and M. Paolone, "Sequential Discrete Kalman Filter for Real-Time State Estimation in Power Distribution Systems: Theory and Implementation," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2358–2370, Sep. 2017, ISSN: 0018-9456, 1557-9662. DOI: 10.1109/TIM.2017.2708278.

[36] R. Madbhavi, H. S. Karimi, B. Natarajan, *et al.*, "Tensor Completion based State Estimation in Distribution Systems," in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, USA: IEEE, Feb. 2020, pp. 1–5, ISBN: 978-1-72813-103-0. DOI: 10.1109/ISGT45199.2020.9087747.

[37] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Enhanced Tensor Completion Based Approaches for State Estimation in Distribution Systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 5938–5947, Sep. 2021, ISSN: 1551-3203, 1941-0050. DOI: 10.1109/TII.2020.3035449.

[38] R. Palaniappan, B. Bauernschmitt, D. Hilbrich, *et al.*, "An Intelligent Measurement and Control Device for Active Distribution Grids," in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Oct. 2020, pp. 975–979. DOI: 10.1109/ISGT-Europe47291.2020.9248862. (visited on 05/07/2025).

[39] R. Nuqui and A. Phadke, "Phasor Measurement Unit Placement Techniques for Complete and Incomplete Observability," *IEEE Transactions on Power Delivery*, vol. 20, no. 4, pp. 2381–2388, Oct. 2005, ISSN: 0885-8977. DOI: 10.1109/TPWRD.2005.855457.

[40] A. Phadke, "Synchronized phasor measurements in power systems," *IEEE Computer Applications in Power*, vol. 6, no. 2, pp. 10–15, Apr. 1993, ISSN: 0895-0156. DOI: 10.1109/67.207465.

[41] A. G. Phadke, J. S. Thorp, and K. J. Karimi, "State Estimlatjon with Phasor Measurements," *IEEE Transactions on Power Systems*, vol. 1, no. 1, pp. 233–238, 1986, ISSN: 0885-8950. DOI: 10.1109/TPWRS.1986.4334878.

[42] J. Vijaychandra, B. R. V. Prasad, V. K. Darapureddi, *et al.*, "A Review of Distribution System State Estimation Methods and Their Applications in Power Systems," *Electronics*, vol. 12, no. 3, p. 603, Jan. 2023, ISSN: 2079-9292. DOI: 10.3390/electronics12030603.

[43] A. Von Meier, E. Stewart, A. McEachern, *et al.*, "Precision Micro-Synchrophasors for Distribution Systems: A Summary of Applications," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2926–2936, Nov. 2017, ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2017.2720543.

[44] B. Pinte, M. Quinlan, and K. Reinhard, "Low voltage micro-phasor measurement unit (micro-pmu)," in *2015 IEEE Power and Energy Conference at Illinois (PECI)*, Champaign, IL, USA: IEEE, Feb. 2015, pp. 1–4, ISBN: 978-1-4799-7949-3. DOI: 10.1109/PECI.2015.7064888.

[45] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed Energy Resources Topology Identification via Graphical Modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2682–2694, Jul. 2017, ISSN: 0885-8950, 1558-0679. DOI: 10.1109/TPWRS.2016.2628876.

[46] K. Dehghanpour, Z. Wang, J. Wang, *et al.*, "A Survey on State Estimation Techniques and Challenges in Smart Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312–2322, Mar. 2019, ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2018.2870600.

[47] European Commission. Directorate General for Energy. and Tractebel Impact., *Benchmarking Smart Metering Deployment in the EU-28: Final Report.* LU: Publications Office, 2020.

[48] A. Angioni, T. Schlösser, F. Ponci, *et al.*, "Impact of Pseudo-Measurements From New Power Profiles on State Estimation in Low-Voltage Grids," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 1, pp. 70–77, Jan. 2016, ISSN: 1557-9662. DOI: 10.1109/TIM.2015.2454673.

[49] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Enhancing Observability in Distribution Grids Using Smart Meter Data," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5953–5961, Nov. 2018, ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2017.2699939.

[50] D. Echternacht, M. Franken, A. Moser, *et al.*, "Smart Area Aachen - in field test of meter placement and state estimation algorithms for distribution grids," in *2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, Oct. 2015, pp. 435–439. DOI: 10.1109/ISGT-LA.2015.7381194.

[51] B. R. Pokhrel, B. Bak-Jensen, and J. R. Pillai, "Integrated Approach for Network Observability and State Estimation in Active Distribution Grid," *Energies*, vol. 12, no. 12, p. 2230, Jan. 2019, ISSN: 1996-1073. DOI: 10.3390/en12122230.

[52] W. Biener, S. Killinger, B. Wille-Haussmann, *et al.*, *Modelling Active and Reactive Power of PV-systems as Input for State Estimation*. Nov. 2014.

[53] D. Falabretti, M. Delfanti, and M. Merlo, "Distribution networks' observability: A novel approach and its experimental test," *Sustainable Energy, Grids and Networks*, vol. 13, pp. 56–65, Mar. 2018, ISSN: 2352-4677. DOI: 10.1016/j.segan.2017.12.006.

[54] IEC 61869-2, *Instrument transformers - Part 2: Additional requirements for current transformers*, 2012.

[55] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10, Madison, WI, USA: Omnipress, Jun. 2010, pp. 807–814, ISBN: 978-1-60558-907-7.

[56] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, JMLR Workshop and Conference Proceedings, Jun. 2011, pp. 315–323.

[57] B. Xu, N. Wang, T. Chen, *et al.*, *Empirical Evaluation of Rectified Activations in Convolutional Network*, Nov. 2015. DOI: 10.48550/arXiv.1505.00853. arXiv: 1505.00853 [cs].

[58] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*, Feb. 2016. DOI: 10.48550/arXiv.1511.07289. arXiv: 1511.07289 [cs].

[59] G. Klambauer, T. Unterthiner, A. Mayr, *et al.*, *Self-Normalizing Neural Networks*, Sep. 2017. DOI: 10.48550/arXiv.1706.02515. arXiv: 1706.02515 [cs].

[60] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, Second edition. O'Reilly, 2019, ISBN: 978-1-4920-3264-9 978-1-4920-3261-8.

[61] F. Hutter, J. Lücke, and L. Schmidt-Thieme, "Beyond Manual Tuning of Hyperparameters," *KI - Künstliche Intelligenz*, vol. 29, no. 4, pp. 329–337, Nov. 2015, ISSN: 1610-1987. DOI: 10.1007/s13218-015-0381-0.

[62] F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., *Automated Machine Learning: Methods, Systems, Challenges* (The Springer Series on Challenges in Machine Learning). Cham: Springer International Publishing, 2019, ISBN: 978-3-030-05317-8 978-3-030-05318-5. DOI: 10.1007/978-3-030-05318-5.

[63] T. Nakagawa, Y. Hayashi, and S. Iwamoto, "Neural network application to state estimation computation," in *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, Seattle, WA, USA: IEEE, 1991, pp. 188–192, ISBN: 978-0-7803-0065-1. DOI: 10.1109/ANN.1991.213480.

[64] Y. Zhao, "An overview of machine learning applications in elastic power systems," *Applied and Computational Engineering*, vol. 95, no. 1, pp. 107–117, Oct. 2024, ISSN: 2755-2721, 2755-273X. DOI: 10.54254/2755-2721/95/2024BJ0064.

[65] J.-H. Menke, N. Bornhorst, and M. Braun, *Distribution System Monitoring for Smart Power Grids with Distributed Generation Using Artificial Neural Networks*, Mar. 2019. arXiv: 1801.04705 [cs].

[66] P. N. P. Barbeiro, J. Krstulovic, H. Teixeira, *et al.*, "State estimation in distribution smart grids using autoencoders," in *2014 IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014)*, Langkawi, Malaysia: IEEE, Mar. 2014, pp. 358–363, ISBN: 978-1-4799-2422-6 978-1-4799-2421-9. DOI: 10.1109/PEOCO.2014.6814454.

[67] Y. Wang, J. Gu, and L. Yuan, "Distribution network state estimation based on attention-enhanced recurrent neural network pseudo-measurement modeling," *Protection and Control of Modern Power Systems*, vol. 8, no. 1, p. 31, Dec. 2023, ISSN: 2367-2617, 2367-0983. DOI: 10.1186/s41601-023-00306-w.

[68] A. S. Zamzam and N. D. Sidiropoulos, "Physics-Aware Neural Networks for Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4347–4356, Nov. 2020. DOI: 10.1109/tpwrs.2020.2988352.

[69] L. Zhang, Gang Wang, G. Wang, *et al.*, "Real-Time Power System State Estimation and Forecasting via Deep Unrolled Neural Networks," *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4069–4077, Aug. 2019. DOI: 10.1109/tsp.2019.2926023.

[70] A. Pareja, G. Domeniconi, J. Chen, *et al.*, *EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs*, Nov. 2019. DOI: 10.48550/arXiv.1902.10191. arXiv: 1902.10191 [cs].

[71] L. Zhao, Y. Song, C. Zhang, *et al.*, "T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020, ISSN: 1524-9050, 1558-0016. DOI: 10.1109/TITS.2019.2935152.

[72] Simon Stock, Simon Stock, Markus Dressel, *et al.*, "Application of Physics-based Graph Convolutional Network in Real-time State Estimation of Under-determined Distribution Grids," Oct. 2022. DOI: 10.1109/isgt-europe54678.2022.9960577.

[73] M. Huang, Z. Wei, G. Sun, *et al.*, "Hybrid State Estimation for Distribution Systems With AMI and SCADA Measurements," *IEEE Access*, vol. 7, pp. 120 350–120 359, 2019, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2937096.

[74] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-Driven Learning-Based Optimization for Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4805, Nov. 2019, ISSN: 0885-8950, 1558-0679. DOI: 10.1109/TPWRS.2019.2909150.

[75] L. Böttcher, H. Wolf, B. Jung, *et al.*, *Solving AC Power Flow with Graph Neural Networks under Realistic Constraints*, Apr. 2022. arXiv: 2204.07000 [cs].

[76] W. Luan, J. Peng, M. Maras, *et al.*, "Smart Meter Data Analytics for Distribution Network Connectivity Verification," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964–1971, Jul. 2015, ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2015.2421304.

[77] G. Cavraro, V. Kekatos, and S. Veeramachaneni, "Voltage Analytics for Power Distribution Network Topology Verification," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 1058–1067, Jan. 2019, ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2017.2758600.

[78] R. Singh, E. Manitsas, B. C. Pal, *et al.*, "A Recursive Bayesian Approach for Identification of Network Configuration Changes in Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1329–1336, Aug. 2010, ISSN: 0885-8950, 1558-0679. DOI: 10.1109/TPWRS.2010.2040294.

[79] B. Hayes, A. Escalera, and M. Prodanovic, "Event-triggered topology identification for state estimation in active distribution networks," in *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Ljubljana, Slovenia: IEEE, Oct. 2016, pp. 1–6, ISBN: 978-1-5090-3358-4. DOI: 10.1109/ISGTEurope.2016.7856295.

[80] V. Miranda, J. Krstulovic, H. Keko, *et al.*, "Reconstructing Missing Data in State Estimation With Autoencoders," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 604–611, May 2012, ISSN: 0885-8950, 1558-0679. DOI: 10.1109/TPWRS.2011.2174810.

[81] D. Singh, J. Pandey, and D. Chauhan, "Topology Identification, Bad Data Processing, and State Estimation Using Fuzzy Pattern Matching," *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1570–1579, Aug. 2005, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2005.852086.

[82] R. Arghandeh, M. Gahr, A. Von Meier, *et al.*, "Topology detection in microgrids with micro-synchrophasors," in *2015 IEEE Power & Energy Society General Meeting*, Denver, CO, USA: IEEE, Jul. 2015, pp. 1–5, ISBN: 978-1-4673-8040-9. DOI: 10.1109/PESGM.2015.7286053.

[83] M. Babakmehr, M. G. Simoes, M. B. Wakin, *et al.*, "Compressive Sensing-Based Topology Identification for Smart Grids," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 532–543, Apr. 2016, ISSN: 1551-3203, 1941-0050. DOI: 10.1109/TII.2016.2520396.

[84] *VDE-AR-N 4100 Anwendungsregel: 2019-04: Technische Regeln für den Anschluss von Kundenanlagen an das Niederspannungsnetz und deren Betrieb*, 2019.

[85] M. Wagler and R. Witzmann, "Effects of asymmetrically connected PV and battery systems on the node voltages and pen-conductor currents in low-voltage grids," *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2246–2249, Oct. 2017, ISSN: 2515-0855. DOI: 10.1049/oap-cired.2017.0315.

[86] H. Früh, K. Rudion, A. von Haken, *et al.*, "Evaluation of a Three-Phase Distribution System State Estimation for Operational Use in a Real Low Voltage Grid," in *The 9th Renewable Power Generation Conference (RPG Dublin Online 2021)*, vol. 2021, Mar. 2021, pp. 125–130. DOI: 10.1049/icp.2021.1396.

[87] G. Cavraro, E. Dall'Anese, J. Comden, *et al.*, "Online State Estimation for Time-Varying Systems," *IEEE Transactions on Automatic Control*, vol. 67, no. 10, pp. 5424–5431, Oct. 2022, ISSN: 1558-2523. DOI: 10.1109/TAC.2021.3120679.

[88] G. Cavraro, J. Comden, E. Dall'Anese, *et al.*, "Real-Time Distribution System State Estimation with Asynchronous Measurements," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3813–3822, 2022, ISSN: 1949-3053. DOI: 10.1109/tsg.2022.3171466.

[89] Y. Yuan, K. Dehghanpour, and Z. Wang, "Mitigating Smart Meter Asynchrony Error Via Multi-Objective Low Rank Matrix Recovery," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4308–4317, Sep. 2021, ISSN: 1949-3061. DOI: 10.1109/TSG.2021.3088835.

[90] B. H. Hassan, A. Narayan, D. Babazadeh, *et al.*, "Performance Assessment of State Estimation in Cyber-Physical Energy Systems," in *2021 IEEE Madrid PowerTech*, Jun. 2021, pp. 1–6. DOI: 10.1109/PowerTech46648.2021.9494760.

[91] M. Fotopoulou, S. Petridis, I. Karachalios, *et al.*, "A Review on Distribution System State Estimation Algorithms," *Applied Sciences*, vol. 12, no. 21, p. 11 073, Nov. 2022, ISSN: 2076-3417. DOI: 10.3390/app122111073.

[92] M. Brand, D. Babazadeh, and S. Lehnhoff, "Trust in Power System State Variables based on Trust in Measurements," in *2021 IEEE Madrid PowerTech*, Jun. 2021, pp. 1–6. DOI: 10.1109/PowerTech46648.2021.9494769.

[93] M. Brand, D. Engel, and S. Lehnhoff, "Assess: Anomaly sensitive state estimation with streaming systems," *Energy Informatics*, vol. 6, no. 1, p. 19, Oct. 2023, ISSN: 2520-8942. DOI: 10.1186/s42162-023-00276-1.

[94] X. Zhou, Z. Liu, Y. Guo, *et al.*, "Gradient-Based Multi-Area Distribution System State Estimation," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5325–5338, Nov. 2020, ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2020.3003897.

[95] S. Ackerman, E. Farchi, O. Raz, *et al.*, *Detection of data drift and outliers affecting machine learning model performance over time*, Sep. 2022. DOI: 10.48550/arXiv.2012.09258. arXiv: 2012.09258 [stat].

[96] B. Rout and B. Natarajan, "Bayesian Deep Unfolding for State Estimation in Power Distribution Grids," in *2024 IEEE Kansas Power and Energy Conference (KPEC)*, Apr. 2024, pp. 1–6. DOI: 10.1109/KPEC61529.2024.10676073.

[97] R. L. Olsen, S. Hassani, T. Pedersen, *et al.*, *A stochastic approach to estimate distribution grid state with confidence regions*, Apr. 2024. DOI: 10.48550/arXiv.2404.15722. arXiv: 2404.15722 [stat].

[98] VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V., "Der Digitale Zwilling in der Netz- und Elektrizitä tswirtschaft," Offenbach am Main, Tech. Rep., 2023.

[99] R. Singh, B. Pal, and R. Vinter, "Measurement placement in distribution system state estimation," in *2009 IEEE Power & Energy Society General Meeting*, Calgary, Canada: IEEE, Jul. 2009, pp. 1–1, ISBN: 978-1-4244-4241-6. DOI: 10.1109/PES.2009.5275899.

[100] J. Hiry, C. Kittl, D. Sen Sarma, *et al.*, *SIMONA - A Discrete-Event Distribution Grid Simulation Environment*, https://github.com/ie3-institute/simona, version 3.0.0, Aug. 7, 2023. [Online]. Available: https://simona.ie3.e-technik.tu-dortmund.de.

[101] S. B. (Destatis), *Stromverbrauch der privaten Haushalte nach HaushaltsgröSS enklassen*. [Online]. Available: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/UGR/private-haushalte/Tabellen/stromverbrauch-haushalte.html (visited on 02/05/2025).

[102] J. Kays, *Agent-Based Simulation Environment for Improving the Planning of Distribution Grids: Agentenbasierte Simulationsumgebung Zur Verbesserung Der Planung von Verteilnetzen* (Reihe Ie3 - Institut Für Energiesysteme, Energieeffizienz Und Energiewirtschaft 14), 1. Aufl. Göttingen: Sierke, 2014, ISBN: 978-3-86844-662-3.

[103] J. Figgener, D. Haberschusz, S. Zumühlen, *et al.*, "Speichermonitoring BW," Tech. Rep., 2021.

[104] N. B. GmbH, *Veröffentlichungen - Netze BW GmbH*, https://www.netze-bw.de/unternehmen/veroeffentlichungen#2-1-6. (visited on 02/06/2025).

[105] T. Hartmann and Verband der Netzbetreiber, Eds., *Lastprofile für unterbrechbare Verbrauchseinrichtungen: LPuVe-Praxisleitfaden*, 1. Ausg. Frankfurt am Main Heidelberg: VWEW-Energieverlag, 2003, ISBN: 978-3-8022-0706-8.

[106] J. Hiry, "Agent-based discrete-event simulation environment for electric power distribution system analysis," Ph.D. dissertation, Düren, 2022, ISBN: 9783844084627.

[107] C. Kittl, "Entwurf und Validierung eines individualitätszentrierten, interdisziplinären Energiesystemsimulators basierend auf ereignisdiskreter Simulation und Agententheorie," Ph.D. dissertation, TU Dortmund, Dortmund, Feb. 2022.

[108] P. Moritz, R. Nishihara, S. Wang, *et al.*, "Ray: A distributed framework for emerging AI applications," in *13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 18)*, Carlsbad, CA: USENIX Association, Oct. 2018, pp. 561–577, ISBN: 978-1-939133-08-3.

[109] L. Li, K. Jamieson, A. Rostamizadeh, *et al.*, *A System for Massively Parallel Hyperparameter Tuning*, Mar. 2020. DOI: 10.48550/arXiv.1810.05934. arXiv: 1810.05934 [cs].

[110] VDEW, *Repraesentative VDEW Lastprofile*, 1999. [Online]. Available: https://www.bdew.de/media/documents/1999_Repraesentative-VDEW-Lastprofile.pdf (visited on 03/11/2025).

[111] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[112] H. Hersbach, B. Bell, P. Berrisford, *et al.*, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, ISSN: 0035-9009, 1477-870X. DOI: 10.1002/qj.3803.

[113] V. FNN, "Standardisiertes Vorgehen für die Durchführung der Netzzustandsermittlung auf Basis von Echtzeit-Messwerten in der Niederspannung zur Einhaltung von Mindestanforderungen an deren Sensitivität und Spezifität," Tech. Rep., 2025. [Online]. Available: https://www.bundesnetzagentur.de/DE/Beschlusskammern/1_GZ/BK6-GZ/2022/BK6-22-300/Mitteilung/Mitteilung_3/VDE_FNN_Empfehlung_zu_Tenorziffer_2e.pdf?___blob=publicationFile&v=1 (visited on 03/12/2025).

[114] Y. Gal and Z. Ghahramani, *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, Oct. 2016. arXiv: 1506.02142 [cs, stat].

List of Abbreviations

AAD	Absolute Average Deviation
DLAE	Deep Linear Autoencoder
DG	Distribution Grid
DSSE	Distribution System State Estimation
FNN	Forum Network Technology and Network Operation
GNN	Graph Neural Network
GPS	Global Positioning System
LSTM	Long Short-Term Memory
LV	Low Voltage
MC	Monte Carlo
ML	Machine Learning
MLP	Multilayer Perceptron
μPMU	Micro-Phasor Measurement Unit
MV	Medium Voltage
ANN	Artificial Neural Network
NSE	Neural State Estimation
NSH	Night Storage Heating
PMU	Phasor Measurement Unit
PSDM	PowerSystemDatamodel
PV	Photovoltaic
RAD	Relative Average Deviation
ReLU	Rectified Linear Unit
RES	Renewable Energy Sources
RMSE	Root Mean Square Error
SCADA	Supervisory Control and Data Acquisition
SE	State Estimation
VDE	Association for Electrical, Electronic & Information Technologies
WLS	Weighted Least Squares

List of Symbols

Mathematical Notation and Operators

$\bar{\cdot}$	Mean Value
$\lfloor \cdot \rfloor$	Floor Operator
$\vec{\cdot}$	Vector Notation

Indices

rated	Rated Value i
a	Angle, as in y_a
anomalous	Anomalous
d	Day Type
gen	Scaled Generation
i	Generic Index
i	Transformer i
k	Bin Index
load	Scaled Load
pseudo	Pseudo
remove	Bus to Remove
s	Season
t	Time Step
threshold	Threshold
valid	Valid
x	Model Input
y	Model Output

Sets

B	Set of Buses
E	Set of Reconstruction Errors
K	Set of All Day Types
S	Set of All Seasons
V	Set of Absolute Voltage Deviations

Symbols

Δ	Deviation

δ	Convergence Criterion
η	Relative Frequency
Λ	Transformer Transformation Ratio
λ	Tap Position
$\mathcal{D}$	DLAE model
$\mathcal{P}$	Percentile
ρ	Pearson Correlation
σ	Standard Deviation
τ	Threshold
θ	Average Ambient Temperature
$\tilde{F}$	Normalized Histogram Bin Frequency
a	Voltage Angle
AAD	Absolute Average Deviation
B	Histogram Bin
b	Histogram Bin Boundary
c	Stopping Criterion Number of Iterations
$C1$	C1 Success Criterion
$C2$	C2 Success Criterion
D	Data Set
d	Dijkstra Distance
dV	Relative Voltage Magnitude Increase
$e_{\%}$	Maximum Error in %
F	Histogram Bin Frequency
G	Number of Forward Passes
M	DSSE Model
MSE	Mean Squared Error
NR	Noise Reduction
o	Observability Rate
P	Rated Power
r	Amount of Buses to Remove
$RMSE$	Root Mean Squared Error
s	Scaling Factor
SR	Success Rate
T	Total Number of Time Steps in Data Set
V	Voltage in kV
v	Voltage in p.u.
w	Weighting Factor
x	Input Signal
y	Model Output

List of Figures

List of Tables

Evidence of Scientific Activity

Scientific Publications

[P1] S. Balduin, E. M. Veith, A. Berezin, S. Lehnhoff, T. OberlieSSen, C. Kittl, J. Hiry, C. Rehtanz, G. Torres-Villarreal, S. Leksawat, A. Kubis, and M.-A. Frankenbach, "Towards a Universally Applicable Neural State Estimation through Transfer Learning," in *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, Espoo, Finland: IEEE, Oct. 2021, pp. 1–6, ISBN: 978-1-66544-875-8. DOI: 10.1109/ISGTEurope52324.2021.9639919.

[P2] J. Hiry, C. Kittl, D. S. Sarma, T. OberlieSSen, and C. Rehtanz, "Multivoltage level distributed backwardforward sweep power flow algorithm in an agent-based discrete-event simulation framework," *Electric Power Systems Research*, vol. 213, p. 108 365, Dec. 2022, ISSN: 03787796. DOI: 10.1016/j.epsr.2022.108365.

[P3] Alexander Berezin, S. Balduin, T. OberlieSSen, S. Peter, E. Veith, and S. Lehnhoff, "Application of Recurrent Graph Convolutional Networks to the Neural State Estimation Problem," *International Journal of Electrical and Electronic Engineering & Telecommunications*, vol. 12, no. 3, pp. 209–215, May 2023.

[P4] S. Balduin, A. Berezin, T. OberlieSSen, and S. Peter, "A training algorithm for neural state estimation using a simulation-as-a-service infrastructure," in *The 37th annual European Simulation and Modelling Conference*, 2023.

[P5] A. Berezin, S. Balduin, T. OberlieSSen, S. Peter, and E. M. Veith, *On zero-shot learning in neural state estimation of power distribution systems*, Aug. 2024. DOI: 10.48550/arXiv.2408.05787. arXiv: 2408.05787 [cs]. (visited on 05/07/2025).

[P6] D. Feismann, S. Peter, J. Bao, T. OberlieSSen, and C. Rehtanz, "Determining the influence of energy management systems on simultaneity factors of distribution grids," in *IET Conference Proceedings*, vol. 2024, Jul. 2024, pp. 896–899. DOI: 10.1049/icp.2024.1993.

[P7] T. OberlieSSen, S. Peter, M.-A. Frankenbach, K. auf der Horst, and C. Rehtanz, "Bridging Synthetic Training and Real-World Application: Applied Simulation-based Neural State Estimation," in *NEIS 2024; Conference on Sustainable Energy Supply and Energy Storage Systems*, Sep. 2024, pp. 205–210. DOI: 10.30420/566464028.

[P8] J. Bao, S. Peter, D. Feismann, T. OberlieSSen, and C. Rehtanz, "Co-Simulation Approach for Optimizing Flexibility Usage in an Agent-Based Discrete-Event Simulation Environment," in *CIRED*, Accepted for publication, Jun. 2025.

[P9] D. Feismann, T. OberlieSSen, U. Häger, and C. Rehtanz, "Evaluation for Necessary Power Demand for Electric Truck Charging on Motorways," in *CIRED*, Accepted for publication, Jun. 2025.

[P10] T. OberlieSSen, D. Feismann, F. Klausmann, O. Felix, and C. Rehtanz, "Simulation-Based Analysis of Grid Impact from Large-Scale HDEV Charging Infrastructure," in *ICCSS*, Accepted for publication, May 2025.

[P11] S. Peter, D. Feismann, J. Bao, T. OberlieSSen, and C. Rehtanz, "Energy management and flexibility quantification in a discrete event distribution grid simulation," in *PowerTech*, Accepted for publication, Jun. 2025.

Supervised Master Theses

[M1] S. Peter, “Entwicklung und Implementierung einer verteilten Scheduling-Strategie für eine ereignisbasierte Verteilnetzenergiesystemsimulation,” Supervisors: J. Hiry, C. Kittl, T. OberlieSSen, Unpublished Master Thesis, Faculty of Computer Science, Chair IV - Practical Computer Science, TU Dortmund University, 2021.

[M2] N. Steffan, “Development and implementation of an agent-based model to assess approaches that efficiently account stochastic charging needs of battery-electric vehicles,” Supervisors: C. Kittl, J. Hiry, T. OberlieSSen, Unpublished Master Thesis, Institute of Energy Systems, Energy Efficiency and Energy Economics, TU Dortmund University, 2021.

[M3] J. Bao, “Entwicklung und Analyse eines stochastischen Modells zur Ersatzmodellierung von Mittelspannungs-Ortsnetzstationen,” Supervisors: T. OberlieSSen, S. Peter, Unpublished Master Thesis, Institute of Energy Systems, Energy Efficiency and Energy Economics, TU Dortmund University, 2022.

[M4] F. Rennebaum, “Durchführung einer Netzzustandsschätzung in Niederspannungsnetzen mit künstlichen neuronalen Netzen,” Supervisors: T. OberlieSSen, T. Schwierz, Unpublished Master Thesis, Faculty of Experimental Physics IV, TU Dortmund University, 2023.

Supervised Bachelor Theses

[B1] V. Zachopoulos, “Konzeptionierung und Implementierung einer Kommunikationsplattform zum Einsatz von SIMONA als “Simulation as a Service",” Supervisors: T. OberlieSSen, C. Kittl, J. Hiry, Unpublished Bachelor Thesis, Faculty of Computer Science, Chair IV - Practical Computer Science, TU Dortmund University, 2021.

[B2] M. Staudt, “Simulation und Analyse von räumlich und zeitlich aufgelösten Interaktionsmustern zwischen individueller Mobilität und elektrischem Verteilnetz,” Supervisors: C. Kittl, T. OberlieSSen, Unpublished Bachelor Thesis, Institute of Energy Systems, Energy Efficiency and Energy Economics, TU Dortmund University, 2022.