Data-Driven Dynamic Modeling of Active Distribution Networks using Artificial Intelligence Techniques

A thesis accepted by the Faculty Energy-, Process- and Bio-Engineering of University of Stuttgart to attain the degree of

Doctor of Engineering (Dr.-Ing.)

submitted by

Georgios Mitrentsis
born in Edessa, Greece

Examiner: Univ.-Prof. Dr.-Ing. Hendrik Lens
Co-examiner: Univ.-Prof. Dr.-Ing. Albert Moser

Date of the examination: 17.02.2023

Institute of Combustion and Power Plant Technology - IFK
University of Stuttgart
2023

To my family.

Declaration of Authorship

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel *Data-Driven Dynamic Modeling of Active Distribution Networks using Artificial Intelligence Techniques* selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe; aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

Declaration of Authorship

I hereby certify that the dissertation entitled *Data-Driven Dynamic Modeling of Active Distribution Networks using Artificial Intelligence Techniques* is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly quoted.

Name/Name: Georgios Mitrentsis

Unterschrift/Signed: 

Datum/Date: 25.09.2022

Abstract

Dynamic load models are fundamental components for simulating and analyzing a power system. Due to their utmost importance, a wide range of dynamic models have been introduced within the past years, able to accurately capture the dynamic behavior of conventional loads. However, new grid technologies, such as distributed generation (DG) and controllable loads, are gradually emerging, transforming the purely load-composed distribution networks to active distribution networks (ADNs). Therefore, it is vital for Transmission System Operators (TSOs) to upgrade the existing load models and reassess the stability of their system.

Motivated by study conducted in Germany several years ago, this thesis attempts to determine the steady-state relationship between the frequency and active power using real measurement data. As a next step, the respective methodology for developing load models, which was proposed several years ago, is further extended so that it can be applied to modern distribution networks with DG. The results of the "old" and the current study are compared revealing how much the load dynamics have changed within the past years.

The thesis continues with introducing a three-stage methodology to effectively build a set of dynamic models for an ADN based on field measurements. In the first stage, the proposed clustering method identifies and then discards all the irrelevant data for the model parameter estimation. In the next stage, the remaining data are clustered into groups with similar dynamics while in the third stage, a nonlinear dynamic model is developed for each of the derived groups. It is concluded that, in spite of the large number of measurements representing a wide range of grid configurations, the general dynamic characteristics of an ADN can be accurately

captured using a limited number of models.

Finally, this thesis aims at modeling the uncertainty induced by the stochastic nature of load and DG. To do so, a new probabilistic dynamic model is proposed that, except for its response, additionally yields the corresponding predictive uncertainty. This probabilistic model is further enhanced so that it can directly incorporate time and weather variables. It is shown that this model can decode the influence of this kind of exogenous variables and convert it into more accurate predictions.

Kurzfassung

Lastmodelle sind grundlegende Komponenten für die Simulation und Analyse eines Stromnetzes. Aufgrund ihrer großen Bedeutung wurde in den letzten Jahren eine große Anzahl dynamischer Modelle eingeführt, die das dynamische Verhalten konventioneller Lasten genau modellieren können. Neue Netzkomponenten, wie z.B. dezentrale Erzeugung und steuerbare Lasten, entstehen jedoch nach und nach und verwandeln lastdominierte Verteilnetze in aktive Verteilnetze. Daher ist es für die Übertragungsnetzbetreiber (ÜNB) unerlässlich, die bestehenden Lastmodelle zu aktualisieren und die Stabilität ihres Systems neu zu bewerten.

Angeregt durch eine vor einigen Jahren in Deutschland durchgeführte Studie wird in dieser Dissertation versucht, die stationäre Beziehung zwischen Frequenz und Wirkleistung zu bestimmen. Im nächsten Schritt wird die entsprechende Methodik zur Entwicklung von Lastmodellen, die vor einigen Jahren vorgeschlagen wurde, so erweitert, dass sie auf moderne Verteilnetze mit dezentraler Erzeugung angewendet werden kann. Die Ergebnisse der "alten" und der aktuellen Studie werden verglichen und zeigen, wie stark sich die Lastdynamik in den letzten Jahren verändert hat.

Im weiteren Verlauf der Dissertation wird eine dreistufige Methodik vorgestellt, um eine Reihe von dynamischen Modellen für ein aktives Verteilnetz zu erstellen. In der ersten Stufe identifiziert die vorgeschlagene Clustermethode alle relevanten Daten für die Parameteridentifikation. In der nächsten Stufe werden diese Daten in Gruppen mit ähnlicher Dynamik geclustert. In der dritten Stufe wird ein nichtlineares dynamisches Modell für jede der generierten Gruppen entwickelt. Man kommt zu dem Schluss, dass trotz der großen Anzahl von Messungen, die ein breites Spek-

trum von Netzkonfigurationen repräsentieren, das allgemeine dynamische Verhalten eines aktiven Verteilnetzes mit einer begrenzten Anzahl von Clustern genau erfasst und modelliert werden kann.

Schließlich zielt diese Dissertation darauf ab, die Unsicherheit zu modellieren, die durch die stochastische Natur der Last und der dezentralen Erzeugung entsteht. Zu diesem Zweck wird ein neues probabilistisches dynamisches Modell vorgeschlagen, das abgesehen von seiner deterministischen Antwort auch die entsprechende Prognoseunsicherheit liefert. Dieses probabilistische Modell wird weiter verbessert, so dass es Zeit- und Wettervariablen direkt einbeziehen kann. Es wird gezeigt, dass dieses Modell den Einfluss dieser Art von exogenen Variablen entschlüsseln und in genauere Vorhersagen umsetzen kann.

Preface

The past three years have been an unforgettable journey. I still recall reading the preface of other dissertations when I first started working at University of Stuttgart. At that time, I was really wondering whether I get the chance to reach a point where I could write the preface of my own thesis. It seemed like a distant dream. And here we are. I am currently writing the last remaining part of my PhD thesis. It may sound cliche but nothing would be possible without the people that helped me and supported me throughout this amazing journey.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Hendrik Lens, for everything he has done for me the past three years. I am so grateful that he considered my application, trusted me, and gave me the opportunity to join his group. He was always there to guide me through my research, challenge my models, motivate me, and discuss any idea that might pop into my head. I feel very lucky to have had a so open-minded, intelligent, and kind supervisor. Finally, I would like to thank him for the support and the freedom that he gave me to explore new research paths in the context of power systems.

Special thanks go to Prof. Albert Moser for devoting his valuable time to evaluate this thesis. Furthermore, I would like to thank TransnetBW and in particular, Dr. Joachim Lehner and Hans Abele, for supporting the research presented in this thesis, as well as Netze BW GmbH and Stuttgart Netze GmbH for providing us the data from their substations.

During the past three years, I had the chance to work and interact with a few great people within our group. I would like to thank Christian Schöll, Benjamin Müller, Michael van der Straeten, Philipp Maucher, Oussama Alaya, Daniel Scheifele, and

Johannes Lips for integrating me so nicely into the team, helping me with the German language, and making my time at IFK so pleasant.

For me, life without good friends does not make sense. I would like to thank the Greek friends of Stuttgart for making me feel like home in Germany. Also, I want to thank Alkis, Billakos, Iraklis, Loukas, Panagiotis, Stelios, and Vangelis for being such great friends and even though we are all spread out across Europe and do not see each other that often anymore, they are always there for me.

I cannot find words to express my gratitude to my parents Giannis and Maria and my little brother Thanasis, who have been unconditionally supporting me throughout my life and always stand by my side. Their contribution to achieving my personal goals is immense. My last thanks and love belong to my future wife Eirini for always believing in me and being a source of joy, happiness, and motivation. This thesis would not have been possible without her constant support.

Contents

List of Figures

List of Tables

Acronyms

ADN Active distribution network.

ANN Artificial neural network.

CDF Cumulative distribution function.

CHP Combined heat and power.

CLM Composite load model.

CRPS Continuous ranked probability score.

DG Distributed generation.

DSO Distribution system operator.

EHV Extreme high voltage.

ERM Exponential recovery model.

EV Electric vehicle.

FD Frequency domain.

FFT Fast Fourier transform.

GP Gaussian process.

HV High voltage.

IM Induction motor.

LSTM Long short-term memory.

MV Medium voltage.

PCC Point of common coupling.

pu per unit.

PV Photovoltaic.

RES Renewable energy sources.

RMSE Root mean square error.

RNN Recurrent neural network.

TD Time domain.

TSO Transmission system operator.

WECC Western Electricity Coordinating Council.

Chapter 1

Introduction

This chapter introduces the problem of dynamic load modeling, which was recently extended to the dynamic modeling of active distribution networks (ADNs). The chapter briefly discusses the main approaches that have been proposed in the literature and their characteristics. Next, current challenges and issues in the existing approaches are identified in order to set the ground for the objectives and the contributions of this work. The chapter ends with the thesis outline and a list of publications derived by this work.

Before starting with the background and motivation of this work, it is important to clarify the terms *dynamic load model* and *equivalent model*, which form the main core of this thesis. Those terms refer to the dynamic models used in power system stability studies in order to capture the dynamic behavior of a distribution network as a whole during electromechanical transients. Those models should not be confused with "load models" that generate time series of load forecasts.

1.1 Background and motivation

In the last two decades, power systems worldwide experience tremendous changes in an attempt to mitigate climate change [5, 6]. Following this trend, the European power system has already changed drastically while even more changes to-

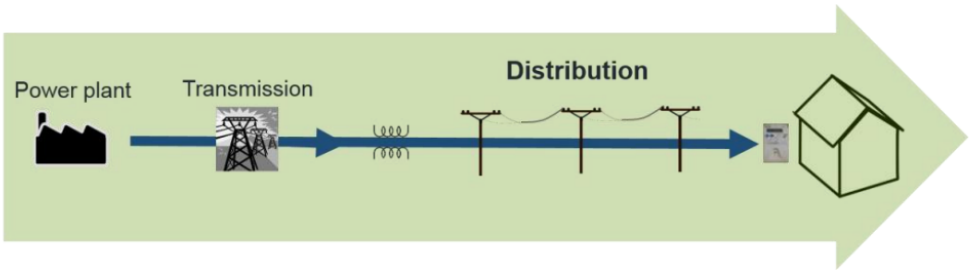


Figure 1.1: Traditional passive distribution network [1].

wards a fossil-fuel-free system are expected within the next years [7]. The climate policies and the technological advances in renewable energy sources (RES) have led to the wide penetration of distributed generation (DG) in distribution networks. Under these new circumstances, the passive load dominated distribution networks (Fig. 1.1) are being transformed to active systems, usually known as "ADNs" in the literature [2, 8].

RES are predominantly connected to the grid through power electronics which do not have the same stabilizing properties as a synchronous generator [5]. At the same time, plug-in electric vehicles (EVs) are typically viewed as a promising pathway to decarbonize the transportation sector in the long-term. Moreover, the penetration of controllable loads comprised by power electronics has significantly increased [2]. As a result, RES, EVs, and the changing nature of loads introduce new variables in the power system equation (Fig. 1.2). Power system stability is becoming a crucial issue for planning and operating transmission systems and dynamic security assessments gain in importance [9]. Hence, this transition requires appropriate tools and changes in power system operation and planning [10].

More importantly, the changing distribution grid architecture yields the necessity of upgrading the existing dynamic models that are still deployed by the Transmission System Operators (TSOs) for estimating the impact of the distribution networks on the overall power system dynamics [11]. This impact can be split into two main components. The first one originates from the load and is also known as the *self-regulating effect*. The loads react to changes in frequency and voltage causing, most

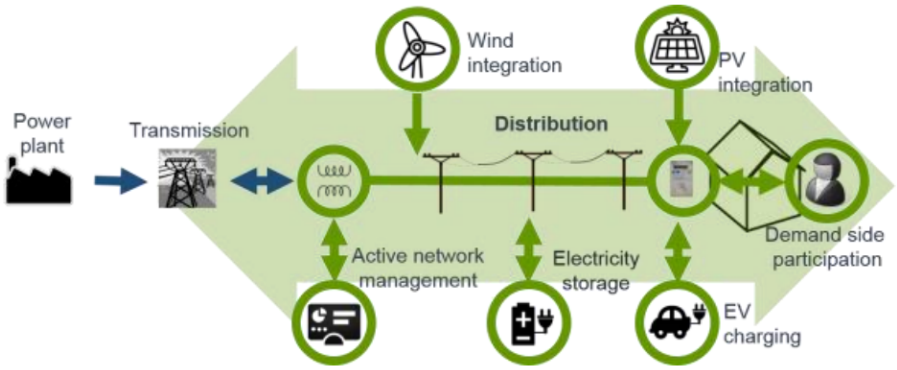


Figure 1.2: Modern active distribution network [1].

of the times, a stabilizing effect on active and reactive power imbalances. The second component is attributed to the DG which may support the voltage and/or the frequency of the system based on the given control strategies [12].

Therefore, it becomes clear that it is vital for TSOs to accurately model the underlying ADNs in order to adequately assess the stability of their system. The necessity of accurate models is also emphasized by the fact that many blackouts have showed the consequences of incorrect modeling of distribution networks in stability studies [13].

To model a power system for stability studies, one solution could be the transmission and distribution system co-simulation (T&D co-simulation), where both systems are modeled in detail. Recent approaches of T&D co-simulation can be found in [14, 15, 16].

However, a detailed model development is a rather delicate task since it demands a concrete knowledge of each individual component in the system, ranging from the various distributed generators, partially with stochastic production, to the exact energy consumption of the power consumers. Consequently, the development of an accurate model which takes into consideration the aforementioned variables is practically impossible. First of all, the computational requirements of such a detailed representation of the system alongside with the time constraints that characterize the dynamic simulations of the power system cannot be met. Secondly, the data

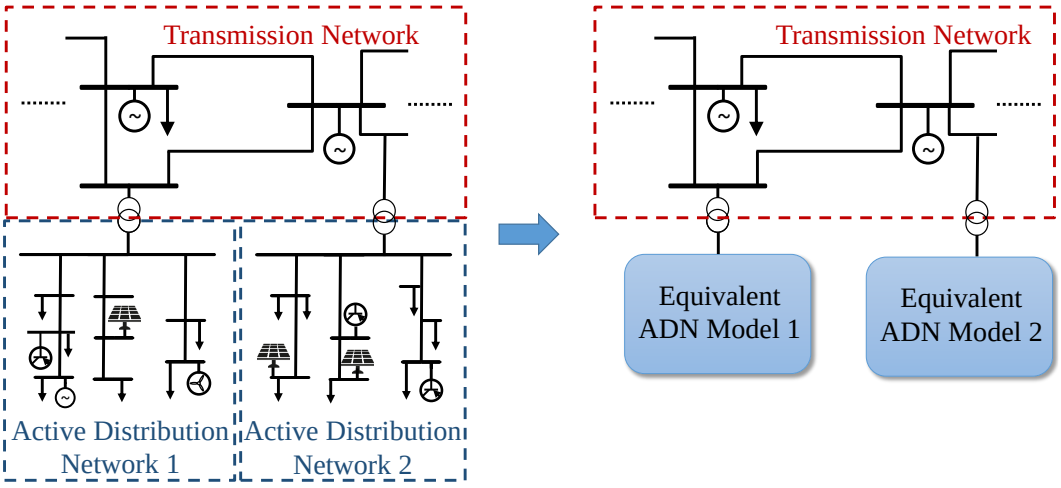


Figure 1.3: Aggregate individual ADN components to develop an equivalent model.

corresponding to load composition such as the total energy consumption of motors, electrical appliances or lighting is unknown [2].

Another solution for modeling distribution networks is the reduced-order or dynamic equivalent model. Both terms refer to the same principle and are used interchangeably in the rest of the thesis. Those reduced-order models are expected to capture the dynamics of a distribution network as a whole without knowing the exact network architecture [17]. This characteristic leads to a significant reduction in model complexity compared to the detailed modeling approach [18]. Nevertheless, those equivalent models need to be characterized by a meaningful trade-off between computational complexity and accuracy. Once this trade-off is adequately met, the dynamic equivalent models can be safely integrated by the TSOs into their models used for stability studies. An illustrative representation of the reduced-order models is shown in Fig. 1.3.

Due to the ability of equivalent models to capture the dynamics of the system while keeping the computational complexity low, significant research effort was put on developing those dynamic models [2].

1.2 Current challenges

Two contrasting methodologies for building dynamic distribution network models, or dynamic load models as usually called, can be found in the literature, namely the component-based [19, 20, 21, 22, 23] and the measurement-based approach [24, 25, 26, 27, 28, 29, 30]. The former relies on detailed knowledge of the components comprising the distribution network and their individual dynamic models whereas the latter is based on field measurements in order to capture the load dynamic behavior. A detailed overview of the existing work based on those two approaches is presented in Chapter 2.

The measurement-based approach has been widely used in the literature since it can be applied to any distribution network without requiring highly detailed knowledge of the system [2]. In this context, the applied measurements should correctly reflect the time-varying dynamics of an ADN. However, the majority of the inspiring research work following the measurement-based approach deploys synthetic data generated by a simulation software or by a small laboratory-scale microgrid [3, 10, 11, 29, 30, 31, 32, 33, 34, 35, 36, 37]. Yet the accordance between the synthetic and real data may be questioned since no validation results have been documented. In reality, ADN dynamics can strongly vary based on geographical and seasonal variations [2]. Both the load and the DG are constantly changing not only in magnitude but also in composition. Thus, models derived from synthetic data may fail to model ADNs in a realistic fashion.

A significant smaller amount of research work has employed real data in order to build dynamic equivalent load models [24, 25, 28, 26, 27, 38, 39]. However, the data were obtained around a decade ago, when distribution networks were governed by passive load components and the measurements reflected only conventional load dynamics. Nowadays, DG has a large influence on the system dynamics due to the various grid codes worldwide for voltage support. Therefore, an ADN may exhibit an entirely different dynamic behavior based on whether the DG units are in operation. As a result, at least two fully distinct models may be required to capture the various dynamics of the same ADN.

More importantly, [25, 26, 39] demonstrated the strong time-varying nature of the load model parameters. In this context, it is commonly recognized that models relying only on a small amount of measurements are unreliable, and thus measurements should be acquired over a reasonable amount of time so that seasonal variations are considered [24, 35, 39]. In this regard, the inspiring work of [40] aims at addressing the issue of limited data by acquiring field measurements for a complete year. However, this approach employs purely static models neglecting the system dynamics while it is restricted to traditional load-composed distribution systems, where no DG dynamics are considered. Furthermore, 384 different models were developed for each substation corresponding to different combinations of half hour intervals within a day (48), the type of the day, i.e., weekend or not (2), and the season (4); a fact that resulted in a large number of nearly identical models. Hence, it becomes clear that a meaningful split of the measurement data in order to build various dynamic models, i.e., developing models for combinations of hour, day, month, high or low temperature, while avoiding redundant models is not straightforward.

Furthermore, although measuring over, e.g., a year can capture a wide range of distribution system configurations (topology, load magnitude and composition), irrelevant measurements can be also recorded as the trigger criteria cannot distinguish the cause of the disturbance. In this regard, since those dynamic models are expected to be deployed in power system dynamic simulations and stability studies, data that do not reflect the response of the ADN to an external disturbance should be discarded from the system identification procedure [2]. For example, faults within the examined grid, wrongly triggered measurements, connection/disconnection of big consumers/DG within the examined grid, highly noisy data and outliers can be additionally recorded. Hence, long measuring periods usually result in big datasets, in which manual data processing would be an extremely time-demanding task. Besides that, identifying a different model for each measurement will simply yield overfitted models, valid only for a specific grid configuration. To this end, the optimal number of dynamic models, which adequately capture the general ADN characteristics while maintaining reliable levels of accuracy is an open question and strongly depends on the recorded dataset.

Another crucial issue that should be pointed out is the direct deployment of the initial power values in various model structures, as thoroughly analyzed in Chapter 2. In modern ADNs, the recorded power at the point of common coupling (PCC), i.e., connection point between transmission and distribution network, expresses the net consumption and does not reflect the real power demand but rather the difference between the real load and the power generated by DG. Therefore, using the raw recorded power values may lead to error-prone models, due to the high volatility of DG, which vary their feed-in depending on the weather conditions and the grid codes for voltage support [41]. At the same time, bi-directional power flows that may occur in the context of ADNs pose even more challenges in the development of reliable ADN models [4].

Lastly, state-of-the-art methods employing the measurement-based approach do not account for the stochastic nature of load and DG, where the consumers continuously alter their power demand and the RES adjust their generation based on weather variations. In the literature, almost all the existing approaches are based on deterministic dynamic models. Those models do not model the uncertainty induced by the volatile RES and fluctuating load nor provide information on whether the yielded responses under previously unseen scenarios should be trusted or not.

To sum up, dynamic models based on simulation-generated synthetic data may not always reflect the underlying dynamics in a realistic fashion. For example, the measurements recorded during the blackout in Western North America in 1996 did not agree with the simulation results based on the Western Electricity Coordinating Council (WECC) models [42]. In this context, there is also a lack of real up-to-date measurement data reflecting the influence of the new load types and DG on ADN dynamics. In addition, data processing techniques have been identified as a field that needs to be further enhanced in order to effectively handle the large amount of data collected by measurement devices [2]. Regarding the modeling itself, the ideal number of dynamic models capturing the time-varying nature of loads and DG while balancing accuracy and generalization capability poses one of the most challenging questions in load/ADN modeling [2, 26, 38]. Moreover, complex models characterized by a large number of parameters are hard to be applied in practice. Hence, a

reasonable trade-off between model complexity and performance is necessary [2]. It should be also highlighted that the direct deployment of the raw power values recorded at the PCC may lead to error-prone models in ADNs with DG [41]. Finally, the output of deterministic dynamic models cannot provide information about the uncertainty induced by load and DG, since a single prediction point is yielded for each time step.

1.3 Objectives and contributions of this thesis

As thoroughly described in the previous section, there are several open questions and issues associated with the development of generic yet accurate ADN models for dynamic simulations. To this end, the objective of this work is to propose a complete methodology to develop dynamic ADN models based on the measurement-based approach while addressing the aforementioned challenges. Specifically, the main objectives of this thesis are listed below:

- The development of a dynamic model that takes as input a voltage signal measured at PCC and generates the active or reactive power response of the underlying ADN. Importantly, this dynamic model aims at capturing electromechanical transients that are typically described by a time resolution within the range of $10^{-2}\mathrm{s}$ - $1\mathrm{s}$ [35]. During those transients, the bus voltage may vary by $\pm 10\%$ from the nominal value [43] and thus, small-signals dynamics are modeled. Note that this work does not cover dynamic models capturing large-signal behavior (e.g., due to short circuits), slow dynamics (e.g.,due to thermal loads), electromagnetic transients, and harmonics.

- The development of a dynamic model that can be integrated into power system stability studies.

- The use of recent field measurements acquired over a reasonable amount of time so that seasonal grid variations (different topology, load magnitude and composition) are considered.

- The development of a data processing technique that can handle a large amount of measurements and identify the relevant data with respect to system identification.
- The development of a methodology to determine the optimal number of dynamic models required to capture the general ADN characteristics while maintaining reliable levels of accuracy.
- The development of a model that can capture the nonlinear load and DG dynamics without being extremely complex. The reasonable model complexity is essential for the practical implementation of the derived models in simulation environments.
- The development of a method that can model the uncertainty induced by the stochastic nature of load and DG.
- The incorporation of a list of exogenous variables in the modeling procedure in order to capture their influence on the time-varying ADN dynamics.

This thesis proposes a range of models and methodologies that are able to address the objectives listed above. In particular, the various contributions of this work can be summarized in the following points:

1. All models and methods developed within this research work are based on real field data acquired over a year in several different substations.
2. Contrary to most of the existing approaches assuming a homogeneous dataset of measurements for building a generic ADN model, this thesis presents an approach that renders this assumption unnecessary by proposing a minimum number of models required to accurately capture the time-varying dynamics of an ADN.
3. Based on the imperative need of the ADN modeling approaches to be robust to bad data [2], this thesis suggests an efficient method to automatically detect outliers and irrelevant measurements within a large dataset of recorded signals.

4. This thesis introduces a novel unsupervised learning method for grouping measurements with similar dynamics into clusters. Importantly, this method can be readily applied to any time-varying dynamical system characterized by distinct sets of dynamics. Note that this approach concerns time-varying dynamical systems where no information is available to identify the different set of dynamics from.

5. Unlike most of the existing models that neglect complex DG control schemes for voltage support, the proposed methodology considers the voltage controller dynamics in a realistic fashion.

6. A performance evaluation of four clustering algorithms of different notion and complexity is conducted in order to address their applicability in clustering dynamic behavior.

7. The proposed models tackle the issue regarding the recorded power values at PCC originated by the generation within an ADN. In this context, those models can accurately capture reverse power flows that may occur after a voltage disturbance.

8. This thesis compares the performance of different dynamic ADN models which were developed using the proposed unsupervised learning methodology.

9. The complete modeling methodology is ADN-agnostic, meaning that it can be effortlessly applied in any given ADN, and does not require detailed information about the system, e.g. load/DG types, DG feed-in at the moment of the disturbance, such as other approaches in the literature.

10. A new application of probabilistic machine learning in the modeling of power systems is introduced.

11. This thesis comments on the feasibility and the challenges of applying probabilistic models in the dynamic modeling of ADNs.

12. This thesis proposes a novel probabilistic dynamic model that encodes the stochastic, nonlinear, and complex nature of an ADN mainly caused by the integration of volatile RES and the time-varying nature of the load.

13. This work introduces for the first time a dynamic load/ADN model that can incorporate a set of exogenous variables in order to generate more accurate predictions.

14. Important nonlinear correlations between ADN dynamics and various time and weather variables are revealed.

15. A method for identifying the separate influence of load and DG within an ADN dynamic response is proposed.

16. A comparison between model parameters estimated more than 20 years ago and now is presented in an attempt to extract information about how load composition changed throughout the years.

1.4 Thesis organization

The rest of this thesis is organized as follows.

Chapter 2: This chapter provides a detailed literature review of the load and, recently, ADN modeling. It starts with the implementation details of the two major methodologies, namely the component-based and the measurement-based approach. The chapter continues with the basic model structures and then moves on to more complex ones. Various state-of-the-art models and parameter estimation techniques are also discussed within this chapter.

Chapter 3: This chapter gives an insight on the measurement procedure followed for the needs of this work, the triggering criteria, the acquired data, the preprocessing techniques applied to the data, and the general characteristics of the measured distribution networks. Additionally, the influence of the various DG technologies and a selection of exogenous variables on the ADN dynamics is investi-

gated. This investigation plays a major part in the development of accurate ADN models as confirmed in the following chapters.

Chapter 4: In this chapter, the frequency and power measurements from three EHV/HV-substations are analyzed in an attempt to derive the steady-state relationship between frequency and active power.

Chapter 5: This chapter encapsulates many of the core contributions of this work. It rigorously describes the proposed unsupervised learning method for clustering dynamic behavior that can be applied to any time-varying dynamical system characterized by distinct sets of dynamics, regardless of the science field. Next, the clustering results using the acquired field data are presented and discussed.

Chapter 6: This chapter presents a linear dynamic model that can be easily integrated to any power system simulation software available in the market. The chapter proposes a new methodology to separate a power response into two different signals. The first one is attributed to the load and the second one is attributed to the generation. In doing so, two different models, i.e., one for load and one for generation, can be derived based on those two signal categories. This chapter ends with a comparison between the dynamic load model parameters estimated several years ago and now.

Chapter 7: This chapter introduces a nonlinear dynamic ADN model based on the well-established exponential recovery model (ERM). In the beginning, the ERM and its modified version for capturing reverse power flow are analyzed. Then, their main issue in modeling modern ADNs is identified and a new model structure alleviating this issue is proposed. The chapter ends with validation results using real measurement data.

Chapter 8: This chapter proposes a new way of modeling ADNs by leveraging the latest advances in probabilistic machine learning. Contrary to almost all the approaches found in the literature using deterministic dynamic models, the proposed method introduces a probabilistic dynamic model based on Gaussian processes (GPs) that yields a probability distribution for each time step. Due to the inherent flexibility of the model structure, two different alternatives are established. The first one uses the autoregressive input and output terms as features whereas

the second one additionally deploys a set of exogenous parameters (apart from the autoregressive terms).

Chapter 9: This chapter discusses the applicability of the proposed models and gives guidelines for their optimal integration in power system simulation packages.

Chapter 10: The last chapter provides a summary of this thesis by highlighting the main findings and conclusions. The chapter ends with suggestions and directions for future work.

1.5 Mathematical notations

In this thesis, the following notation style is adopted:

- Matrices and vectors are denoted by bold letters.
- All vectors are column vectors.

1.6 Publication list

Based on the work performed for this thesis, the following publications have been derived:

1. **G. Mitrentsis** and H. Lens, "A Gaussian Process Framework for the Probabilistic Dynamic Modeling of Active Distribution Networks Using Exogenous Variables, Electric Power Systems Research, 2022.
2. **G. Mitrentsis** and H. Lens, "Probabilistic Dynamic Model of Active Distribution Networks Using Gaussian Processes," IEEE PowerTech 2021, Madrid, Spain, 2021.
3. **G. Mitrentsis** and H. Lens, "Data-Driven Dynamic Models of Active Distribution Networks using Unsupervised Learning Techniques on Field Measurements," IEEE Transactions on Smart Grid, 2021.

4. **G. Mitrentsis** and H. Lens, "A Dynamic Active Distribution Network Equivalent for Enhancing the Generalization Capability of the Exponential Recovery Model in Stability Studies," IEEE Transactions on Power Systems, 2021.

5. **G. Mitrentsis** and H. Lens, "Unsupervised learning method for clustering dynamic behavior in the context of power systems," IFAC-PapersOnLine. 2020.

6. **G. Mitrentsis** and H. Lens, "Dynamic modeling of active distribution networks using cluster analysis of field measurement data," NEIS 2020, Hamburg, Germany, 2020.

7. H. Lens, **G. Mitrentsis**, H. Abele, J. Lehner, and C .Schöll, "Measurement data-based identification of the contribution of distribution networks to the self-regulating effect," ETG/GMA Fachtagung Netzregelung und Systemführung, Leipzig, Germany, 2022

In addition, the following publications were motivated by this work, albeit not directly connected with the topic of this thesis.

1. **G. Mitrentsis** and H. Lens, "An Interpretable Probabilistic Model for Short-Term Solar Power Forecasting Using Natural Gradient Boosting," Applied Energy, 2022.

2. **G. Mitrentsis**, M. Liu, and H. Lens, "Open Source Tool for Probabilistic Short-Term PV and Wind Power Forecasting," IEEE PMAPS 2022, Manchester, UK, 2022.

3. **G. Mitrentsis** and H. Lens, "Enabling Energy Transition: An Interpretable AI Model for Probabilistic Solar Power Forecasting," ET. Energiewirtschaftliche Tagesfragen, 2022.

1.7 Supervised master thesis

While working on the topic of this dissertation, the following master thesis were supervised at University of Stuttgart.

1. MengLing Liu - “Wind and solar power generation forecasting using probabilistic machine learning”

2. Samuel Wiertz - “Data-driven modeling of active distribution networks for dynamic simulations”

3. Philip Grant - “Development of a wind and solar generation forecast model using machine learning algorithms”

Chapter 2

Dynamic Load/ADN Modeling - Literature Review

This chapter provides a detailed literature review of the dynamic load and ADN modeling. It starts with the implementation details of the two major methodologies, namely the component-based and the measurement-based approach. The chapter continues with the basic model structures and then moves on to more complex ones. Various state-of-the-art models and parameter estimation techniques are also discussed within this chapter.

2.1 Component-based approach

In the component-based approach, load is generally represented as an aggregation of an industrial, a residential and a commercial load class, which are also divided into different load components, such as refrigeration, electronics, lighting, and cooling/heat pump [22]. A static or a dynamic model can be deployed for the modeling of each one of the load components, while their parameters are extracted through laboratory experiments [44, 45, 46]. Therefore, although this approach does not require the installation of special equipment to perform the field measurements, it cannot be easily implemented in real systems, since it requires a highly detailed load

composition knowledge. At the same time, this approach is characterized by a low flexibility in integrating new loads [2]. The hierarchical structure of the component-based approach is illustrated in Fig. 2.1. For more information, the interested reader is referred to the component-based implementations of [19, 20, 21, 23].

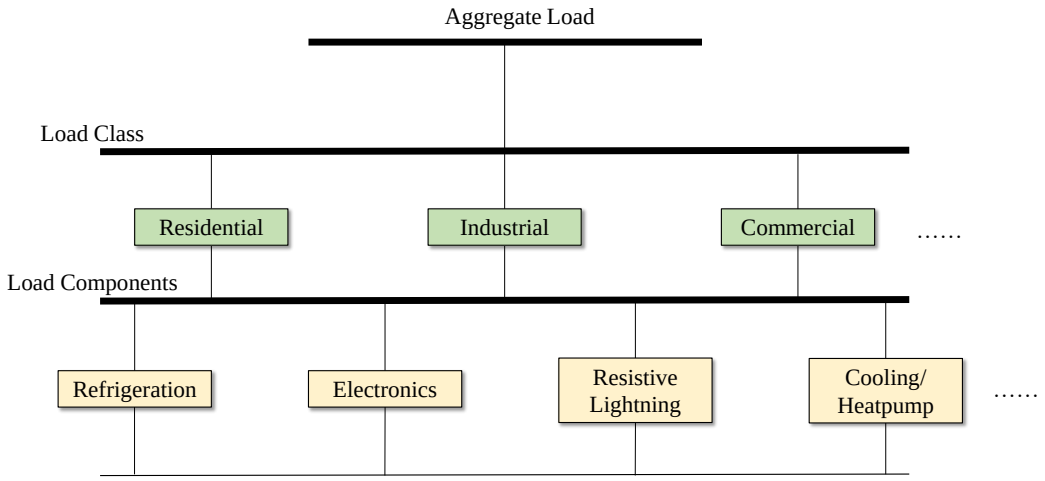


Figure 2.1: Hierarchical structure of the component-based approach [2].

2.2 Measurement-based approach

In the measurement-based approach, the development of the dynamic model can be considered as a system identification task, where an adequate model structure must be initially determined and then, the system parameters need to be estimated based mainly on measurement data. Those data can be either generated by a simulation software (synthetic data) or by on-site measurement devices (real data)[1]. Regarding the model structure, various approaches have been proposed in the literature. Those models describe the mathematical equations that model the relationship between voltage (and rarely frequency) with the active and reactive power at the point of

[1]The terms “real data” and “field data” are used interchangeably referring to the data acquired by on-site measurement devices.

common coupling (PCC). The most widely used model structures are analytically described later in this section.

Once the measurements have been recorded and a specific model structure has been selected, the parameters are estimated by minimizing the square difference between the recorded measurements of active or reactive power and the ones generated by the model. In its simplest form, this cost-minimization problem can be formulated using the following equation:

$$\min \frac{1}{M} \sum_{i=1}^{M} (y_i - \hat{y}_i)^2 \tag{2.1}$$

where y_i corresponds to the measurement values of active or reactive power, $\hat{y}_i$ refers to the model response, and $i = \{1, ..., M\}$ is the measurement index.

As a general rule, voltage magnitudes (V) and rarely frequency (f) are the model inputs while active and reactive power (P, Q) are the model outputs. Load models can be split into static and dynamic loads based on the mathematical equations that are deployed.

2.2.1 Static models

A static load model is usually represented by algebraic equations. These models mainly represent the sensitivity of load active and reactive power to the voltage and frequency changes.

ZIP model

A widely used load model both for steady-state and dynamic studies is the ZIP model, which comprises a polynomial equation with a constant power, a constant current and a constant impedance term [47]:

$$P = P_0 \Big[a_1 \Big(\frac{V}{V_0} \Big)^2 + a_2 \Big(\frac{V}{V_0} \Big) + a_3 \Big], \tag{2.2}$$

$$Q = Q_0 \Big[a_4 \Big(\frac{V}{V_0} \Big)^2 + a_5 \Big(\frac{V}{V_0} \Big) + a_6 \Big], \tag{2.3}$$

where P_0, Q_0, V_0 are the active, reactive power and voltage initial conditions, respectively, and a_i corresponds to the contribution of each term. Representative works that deploy ZIP models and propose parameter estimation techniques for its terms can be found in [13, 24, 48, 49, 50, 51, 52, 53, 54, 55, 56].

Exponential model

The exponential load model is a variation of the ZIP model, where the relationship between power and voltage can be described by an exponential equation [47]:

$$P = P_0\Big(\frac{V}{V_0}\Big)^{n_p}, \tag{2.4}$$

$$Q = Q_0\Big(\frac{V}{V_0}\Big)^{n_q}, \tag{2.5}$$

where n_p and n_q are the voltage dependent parameters of the active and reactive power, respectively. Representative works concerning the static exponential model are included in [40, 48, 57, 58, 59].

Frequency dependent model

This model is generated by simply multiplying either the ZIP or the exponential model equations with the frequency factor

$$\eta_f = \left[1 + a_f(f - f_0)\right], \tag{2.6}$$

where a_f is a frequency sensitivity parameter and f_0 is the nominal frequency of the examined distribution network [47]. In the rest of the chapter, η_{fp} and η_{fq} will refer to the frequency dependent factor of active and reactive power, respectively. In this context, indicative examples of applying the frequency dependent factor are found in [50, 60, 61, 62].

EPRI model

This model was developed by the Electric Power Research Institute (EPRI) and is a combination of the ZIP, Exponential, and frequency dependent models described

above [63]. The active and reactive power are calculated using the formulas

$$P = P_0 \left[\left(P_{a1} \left(\frac{V}{V_0} \right)^{np1} \right) \cdot \eta_{fp} + (1 - P_{a1}) \left(\frac{V}{V_0} \right)^{np2} \right], \tag{2.7}$$

$$Q = P_0 \left[\left(Q_{a1} \left(\frac{V}{V_0} \right)^{nq1} \right) \cdot \eta_{fq1} + \left(\left(\frac{Q_0}{P_0} - Q_{a1} \right) \left(\frac{V}{V_0} \right)^{nq2} \right) \cdot \eta_{fq2} \right], \tag{2.8}$$

respectively. Those two formulas are composed by two terms. In the case of active power, there is a frequency dependent term multiplied by the frequency dependent fraction of active power P_{a1} and a frequency independent term multiplied by the remaining fraction of active power. In the case of reactive power, the first term corresponds to the uncompensated reactive power of the load whereas the second term corresponds to the compensated reactive power; both are multiplied by different frequency dependent factors η_{fq1} and η_{fq2}. Q_{a1} denotes the ration between the uncompensated reactive power of the load to the initial active power consumption. Two representative examples of EPRI model are included in the work of [63, 64].

2.2.2 Dynamic models

In contrast with the static models, dynamic models offer a more accurate representation of the load by incorporating both voltage and time into their active and reactive power formulas. Two of the most widely used dynamic models are the induction motor (IM) model and the exponential recovery model (ERM) [18]. Note that the IM model is usually combined with static models, e.g., ZIP or exponential model. This "hybrid" model is commonly known as composite load model (CLM) [2]. In addition, artificial neural networks (ANNs) have been also proposed as an alternative for dynamic load modeling. Recently, new dynamic model structures have been emerged considering the effects of both load and DG on the system dynamics. Those model structures are typically called ADN models.

Induction motor (IM) model

This model is mainly based on a system of differential equation derived by the equivalent induction motor (IM) circuit shown in Fig. 2.2. Various approaches have been introduced in the literature [65, 66]. They differ with respect to the level of detail and, thus, to the number of parameters of the IM model. In Fig. 2.2, R_s and R_r denote the stator winding and rotor resistance, respectively, while X_s, X_m, and X_r indicate the stator, magnetizing, and rotor reactance, respectively. The rotor slip is denoted with s.

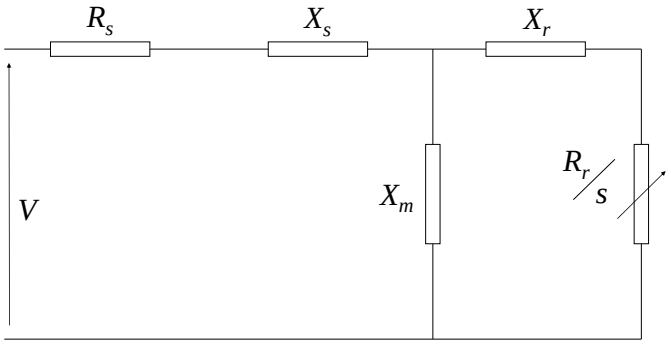


Figure 2.2: Equivalent circuit of IM.

Exponential recovery model (ERM)

This model has produced remarkable results in cases of step voltage jumps caused by on-load tap changing events [67]. The power y (active or reactive) is derived from the nonlinear equations

$$y = y_s - (y_s - y_t) \cdot e^{-\frac{t-t_0}{T_y}}, \tag{2.9}$$

$$y_s = y_0 \left(\frac{V}{V_0}\right)^{N_s}, \tag{2.10}$$

$$y_t = y_0 \left(\frac{V}{V_0}\right)^{N_t}, \tag{2.11}$$

where t_0 is the moment of voltage disturbance, T_y is the time constant of the exponential power recovery, y_s, and y_t are the steady-state and the transient load responses, which are modeled using the exponents N_s and N_t, respectively. Indicative approaches exploiting the ERM have been proposed in [25, 29, 31, 39, 67, 68, 69]. Modifications of this model have been recently introduced in the literature in order to make it suitable for modeling ADNs [4, 41]. A thorough analysis of those two ERM modifications is presented in Chapter 7.

Composite load model

As mentioned, a composite load model is the combination of a static and a dynamic load model. One of the most commonly used composite models is the combined ZIP and IM model [2, 18]. The equivalent circuit of ZIP + IM model is shown in Fig. 2.3. Representative works deploying the composite load model can be found in [24, 25, 38, 50, 60, 66, 70, 71].

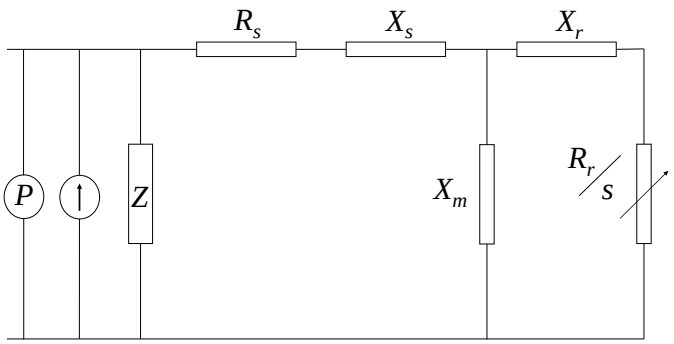


Figure 2.3: Equivalent circuit of ZIP + IM model.

Artificial neural network (ANN) model

Artificial neural networks (ANNs) comprise different layers of processing units, called nodes, which are interconnected by different weights. ANNs have shown very good results in learning how to map input into output data. To do so, an ANN

is typically trained using a procedure called back propagation, which sequentially minimizes the error between the real and model output by adjusting the network parameters. A few indicative papers using ANNs for load modeling can be found in [72, 73, 74]. A more advanced ANN architecture are the recurrent neural networks (RNNs), which are particularly advantageous at modeling sequential data [75]. Due to their advantages, RNNs have been also proposed in load modeling. An inspiring work, which leverages a long short-term memory (LSTM) RNN architecture, has been recently introduced in [3].

Active distribution network (ADN) model

As already described in Chapter 1, an ADN is a distribution network that may contain DG, modern load types, e.g., power electronics, demand management schemes, etc. In this context, a detailed modeling of the individual components comprising an ADN is a computationally demanding task, while the lack of information about those components make the problem even more challenging. To this end, recent approaches have been proposed in order to model an ADN as a whole [2]. Those approaches can be divided into black-box an grey-box models. Black-box models have no physics-based structure and can take any form that maps the input data into the desired output. On the contrary, grey-box models rely on physics-based formulas as well as on physical insight about the examined system. On the one hand, black-box models offer high flexibility on the development of dynamic models, especially in cases of complex ADN architectures. On the other hand, grey-box models can be readily interpreted since they are based on physical relationships.

In the context of black-box approaches, [76] proposes an ARX and a state-space model in order to model a distribution network with generating units. Reference [77] introduces an equivalent black-box model for microgrids using Prony analysis. Included also in the ANN category, [3] proposes the use of LSTMs for modeling dynamical systems. This method is validated using ADNs as an example. Following a similar principle, [78] proposes an equivalent model based on RNNs for modeling a distribution system, which contains a large number of generating units.

Furthermore, linear state-space models are developed for the dynamic modeling of ADNs in the work of [79].

Regarding the grey-box models, a seventh order nonlinear state-space model is introduced in [10]. In particular, this grey-box model comprises a composite load model for modeling the consumers and a synchronous generator connected through a converter to the grid. A extensive evaluation of this model is presented in [32]. A similar approach, albeit static, can be found in [80], where the load is modeled using a ZIP model and the influence of DG is captured by a simple PV model. A more advanced grey-box model was recently introduced in [11, 81]. This equivalent grey-box model is composed by a static exponential model in parallel with an induction motor and an inverter-based generator equipped with modern control capabilities. For instance, controls for low voltage ride-through, reactive current injections for voltage support, and current limitations are included among others.

2.2.3 Parameter estimation

The previous section focuses on the different model structures that have been mostly used in the literature for load/ADN modeling. However, one of the most crucial factors differentiating the various studies presented throughout the years is the deployed methodology for estimating the model parameters. In addition, the data processing techniques applied before the parameter estimation procedure plays also a major role in the development of accurate and generic load/ADN models. This is attributed to the fact that the measurement-based approach requires a reasonable amount of measurement data that should be cleaned, probably clustered, and generally transformed into a suitable form for the system identification [2, 12, 18].

For the parameter estimation, algorithms such as least-squares, genetic algorithms, vector fitting, gradient-based techniques, Levenberg-Marquardt are widely deployed [2]. Examples of the classical least-squares technique and its variations can be found in [82, 83, 84, 85]. Moreover, the study of [86] employ a genetic algorithm, while in [38], a genetic algorithm is combined with the Simplex search method. Similarly, the work of [66] combines a genetic algorithm with the Lev-

enberg–Marquard algorithm whereas the latter is also exploited in [66, 87]. The authors of [29] propose the vector fitting technique as an alternative for estimating the transfer function parameters of the ERM, while gradient-based methods are used in [88, 89, 90].

Recent approaches leverage deep learning techniques in order to estimate the time-varying parameters of the composite load model (CLM). Specifically, [91] deploys a deep generative architecture composed by LSTMs alongside with an autoencoeder for yielding probability density functions of the CLM parameters. In [92, 93], deep reinforcement learning algorithms, such as double deep Q-learning network and imitation and transfer Q-learning, are employed, respectively, where the model parameters are viewed as the state that the respective agent tries to reach. Finally, a multi-modal LSTM method is introduced in [94], while [35] proposes a robust time-varying parameter identification technique. Finally, the work of [31] and [37] exploit artificial neural networks in order to generate a new set of model parameters depending on the grid state, i.e., load and DG mix.

2.3 System reduction based approaches

Apart from the component and the measurement-based approach, a third kind of approach may be also included, even though it is usually omitted from the respective list. For the sake of completeness, it is briefly mentioned in this section. The third approach is based on system reduction techniques and can be split into two main categories, namely the modal analysis and the coherency based methods [95]. The most important work of those two methods is listed in the next two paragraphs.

Modal analysis methods reduce the model order by linearizing around a steady-state point and then neglecting the non-dominant modes [96]. Different modal analysis techniques have been proposed in the literature based on modal truncation [97, 98, 99, 100], Singular Value Decomposition (SVD) [101], Hankel-norm approximations [102], and simple calculations of the dominant poles and the associated residues [96]. However, the modal analysis approaches may not be able to

capture complex power system dynamics, are computationally heavy, and vary a lot in model structure rendering their integration in power system simulation software challenging. Therefore, they have not been widely deployed [95].

On the contrary, coherency-based methods enjoy more popularity compared to modal analysis ones. Those methods attempt to identify similar dynamic properties between generators, which are grouped based on those properties and eventually replaced by one equivalent machine [103]. Important work of coherency-based methods can be found in [103, 104, 105, 106, 107, 108]. Nevertheless, coherency-based methods require knowledge of the network structure as well as the parameters of the underlying components. This information is often unavailable and thus, coherency-based methods may not be applicable for ADN model reduction in practice [95].

2.4 Summary

In this chapter, the two main approaches for building load/ADN models have been described and analyzed. Based on the inherent advantages of the measurement-based approach, more emphasis was put on it. The most commonly used model structures have been identified, split into groups, and presented in detail. Furthermore, various parameter estimation techniques and their application in load/ADN modeling have been discussed.

To conclude, accurate distribution grid modeling is an ongoing research task due to the lack of knowledge regarding load and generation composition. Furthermore, the stochastic behavior of consumers, which is highly influenced by time and weather conditions, hinders the modeling procedure in practice. Active distribution networks comprising generation, storage systems, electric vehicle charging and controlled loads have changed the overall grid characteristics and can remarkably affect system balance and stability. To this end, the existing models (ZIP, ZIP + IM, ERM, etc.) may not be able to accurately mimic the dynamic behavior of active distribution networks and thus, more sophisticated approaches should be introduced.

Additionally, all the aforementioned models implicitly assume that the data deployed for parameter estimation represent qualitatively similar dynamics. In reality however, the active or reactive power controllers of DG may be either active or inactive, depending mainly on whether the DG unit is in operation or not. Hence, utterly different dynamics may be observed based on the DG scheduling. As for the deep learning approaches, they deploy a CLM as model structure, which comprises a static (ZIP) and a dynamic (induction motor) load part. Hence, a CLM cannot fully capture the dynamics of an ADN including, among others, power electronic loads and DG [2]. Moreover, the aforementioned deep learning literature focuses on delivering a highly accurate alternative for nonlinear parameter estimation of complex model structures, whose parameter estimation forms a rather challenging problem. Finally, there are approaches that require the knowledge of load and generation mix in order to generate a set of model parameters. As already pointed out, this information is often hard to obtain even for the respective distribution system operators (DSOs) and thus, those models cannot be implemented in practice.

Chapter 3

Measurement Data and Distribution Networks Under Study

In order to build dynamic load/ADN models based on the measurement-based approach, measurement data are required. This chapter gives an insight on the measurement procedure followed for the needs of this research, the triggering criteria, the acquired data, the preprocessing techniques applied to the data, and the general characteristics of the measured distribution networks. Additionally, the influence of the various DG technologies and a list of exogenous variables on the ADN dynamics is investigated. This investigation plays a major part in the development of accurate ADN models as observed in the following chapters.

3.1 Data acquisition

Generally, the response of the system after a disturbance is required in order to apply system identification techniques. Due to the different nature of voltage and frequency disturbances in a power system, different measuring approaches are followed with respect to those two variables. In this section, the measurement requirements are identified, both for voltage and frequency events. Moreover, information about the hardware used for this study is given.

3.1.1 Measurement requirements and settings

As pointed out in the previous chapters, the dynamic ADN models proposed in this work are based on the measurement-based approach. To apply this method, measurement data of voltage, frequency (rarely in the literature), active, and reactive power are needed. Furthermore, measuring over a year is of utmost importance in order to capture the time-varying nature of loads and DG, as highlighted in Chapter 1.

Since voltage and frequency are likely to experience utterly different disturbances or events, different requirements are set, respectively. In particular, voltage events may be mainly caused by on-load tap changer movements or rarely by short circuits in transmission system [43], whereas frequency events may be initiated by power plant outages or other severe grid faults [109]. In both cases, the RMS values of the following signals are to be recorded:

- the total active and reactive power in all three phases,
- the average of the line-to-line voltages of all three phases,
- the frequency of a reference phase.

Additionally, the exact date and time of each recording needs to be also saved.

Voltage events

Based on [43], voltage changes between 0.5%-2% are sufficient for the parameter estimation of small-signal dynamic load models. However, a possible 0.5% threshold might trigger the measurement unit unnecessarily often due to measurement noise or other irrelevant reasons for the system identification. Therefore, a voltage change threshold of 1% is selected in this study in order to capture the important disturbances while decreasing the probability of recording irrelevant events. In total, six different HV/MV substations were equipped with measurement units. Those substations were deliberately selected so that different load types and DG technologies are considered. More information on these is provided in Section 3.2.

As for the time resolution and duration, the scope of this work is to develop dynamic models for electromechanical transients. Those transient are typically described by a time resolution within the range of 10^{-2}s - 1s [35]. Therefore, the sampling frequency was set to $100\,\mathrm{Hz}$. In addition, a total recording duration of maximum $4\,\mathrm{s}$ seems to be a very common choice in the literature, even for large disturbances [11]. In this study, a rather conservative approach was followed, where each measurement stores voltage, active, and reactive power data for $1\,\mathrm{s}$ before and $9\,\mathrm{s}$ after the voltage disturbance in order to robustly capture the corresponding transient without jeopardizing any loss of information. Therefore, each measurement covers a time period of $10\,\mathrm{s}$, in which the disturbance occurs at the time step $t_d = 1s$. Since no significant variation was observed in the recorded signals during the pre- and after disturbance steady-states, a time horizon of $3.5\,\mathrm{s}$ is deployed in the following sections of this thesis for a visually better presentation of the acquired data. Here, it should be mentioned that PMUs with a minimum sampling frequency of $20\,\mathrm{Hz}$ may also serve as an adequate measuring device, since stability studies are usually performed using a time step of $50\,\mathrm{ms}$ (=$20\,\mathrm{Hz}$) [43].

Frequency events

Contrary to voltage, frequency does not experience step changes, even in cases of large active power imbalances. Furthermore, frequency deviations that may occur throughout a day do not suffice for the dynamic modeling of the system. Under those conditions, long measuring periods are required in an attempt to wait for power plant outages or large grid faults that have an observable impact on the grid frequency. Therefore, two different types of measurements were selected for the frequency events. The first one is similar to the voltage events, where the measurement device is triggered each time there is a frequency change bigger than a user-defined threshold. The second type of measurements are the continuous ones, where the measurement unit records constantly with a time resolution of $2\,\mathrm{s}$. It is worth mentioning that the triggering frequency threshold was modified many times in the beginning of the measurement period, since it was too sensitive to measure-

ment noise. To increase the signal-to-noise ratio, the units for the frequency events were installed in three EHV/HV substations in order to measure a large grid ("network group").

3.1.2 Hardware

The hardware used in this study was selected by the two core project partners, namely the TSO *TransnetBW* and the DSO *NetzeBW*. Two different models of measurement units were deployed based on the type of substation. Specifically, EHV/HV substations were equipped with the unit *PQI-DA smart*, while HV/MV substations with *PQ-Box 150*. Both measurement units are manufactured by *A. Eberle GmbH & Co. KG*. The former were fixedly installed in the corresponding substations whereas the latter were installed for a year in three HV/MV substations and then, they were moved to the other three HV/MV substations for a year, too.

Both types of units have the following important characteristics:

- They are certified for EHV/HV and HV/MV substations and be readily connected to the secondary side of a transformer without the need of major structural changes.
- They offer high sampling frequency ($100\,\mathrm{Hz}$ for RMS measurements).
- They can be programmed to be triggered under desirable voltage and frequency scenarios.
- They can save measurement data internally for a reasonable amount of time ranging from several weeks to a few months.

More information about those units can be found in the respective manufacturer's manuals [110, 111].

3.2 ADNs under study

All ADNs under study are located in southwest Germany, and in particular in the federal state of Baden-Württemberg. The transmission system of this state is mostly controlled by *TransnetBW* (TSO). Due to confidentiality reasons, the name and location of the examined substations cannot be published in this thesis. Therefore, each substation is denoted by a capital letter and each of its connected transformers are represented by a number.

3.2.1 EHV/HV substations

TransnetBW is responsible for the EHV/HV substations of this study. The nominal values of the transformers powering this large "network group" are shown in Table 3.1. The capital letters *M*, *P*, and *W* correspond to each of the three EHV/HV substations, while the numbers (1, 2, ..., 5) after the substation letter (*M*, *P*, and *W*) denote the index of each measured transformer.

Table 3.1: Nominal values of the transformers powering the large network group.

ADN ID	Rated Voltage (kV)	Rated Power (MVA)
Sub.M.1	380/110	300
Sub.M.2	380/110	300
Sub.M.3	380/110	300
Sub.P.1	380/110	300
Sub.P.2	380/110	300
Sub.P.3	380/110	300
Sub.P.4	380/110	300
Sub.P.5	380/110	210
Sub.W.1	380/110	300
Sub.W.2	380/110	300
Sub.W.3	380/110	250

Another important aspect of measuring EHV/HV substations was to determine

which transformers are actually powering the examined network group during the frequency events, as the topology can be changed so that the transformers feed other distribution networks. To do so, the history of the various switching protocols was provided by the TSO in order to reconstruct which transformers fed this group at any given time.

3.2.2 HV/MV substations

As for HV/MV substations, four substations are operated by *NetzeBW* (DSO) while the rest two are operated by *Stuttgart Netze* (DSO). At each of those six substations, two transformers were equipped with measurement devices, which were programmed to record voltage events following the requirements listed earlier in this chapter. Similarly to the EHV/HV substations, each HV/MV substation is given

Table 3.2: Nominal values of the transformers connecting the examined ADNs with the transmission system and number of measurements acquired.

ADN ID	**Rated Voltage (kV)**	**Rated Power (MVA)**	**Number of measurements**
Sub.A.1	110/20	40	492
Sub.A.2	110/20	40	1671
Sub.B.1	110/20	25	300
Sub.B.2	110/20	25	1371
Sub.C.1	110/10	40	1265
Sub.C.2	110/10	40	1365
Sub.D.1	110/10	40	716
Sub.D.2	110/10	40	1027
Sub.E.1	110/20	25	3010
Sub.E.2	110/20	25	1800
Sub.F.1	110/20	25	1796
Sub.F.2	110/20	25	1959

Table 3.3: Nominal values of the installed DG capacity in the ADNs under study.

ADN ID	**Installed DG (MVA)**
Sub.A.1	Wind: 30.8, PV: 16, Other: 3.7
Sub.A.2	PV: 6.4, Other: 0.1
Sub.B.1	Wind: 11.7, PV: 16.4
Sub.B.2	Wind: 16.3, PV: 14.3, Biogas: 3
Sub.C.1	CHP: 2
Sub.C.2	-
Sub.D.1	CHP: 2.7
Sub.D.2	-
Sub.E.1	Wind: 8.6, PV: 8, Other: 0.1
Sub.E.2	Wind: 5.9, PV: 11.6, Other: 0.9
Sub.F.1	PV: 8.5, Biogas: 0.3
Sub.F.2	PV: 10.5

one of the following letters $\{A,B,C,D,E,F\}$, while the indices 1 and 2 are used to indicate the measured transformer at each substation.

The nominal values of the transformers powering each ADN under study and their installed DG capacity are shown in Table 3.2 and 3.3. As observed in those tables, the ADNs differ significantly from each other with respect to the DG technologies and their installed capacities. Moreover, substations A, B, E, and F power rural regions whereas substations C and D power urban ones. It was also communicated by the respective DSOs that the transformers of the same substations supply different consumers with electricity. Therefore, each of the ADNs under study is characterized by a distinct load and DG composition; a fact that is of utmost importance for the development of generic dynamic models.

3.2.3 Timeline of measurements

The exact time period that each substation was measured is illustrated in Fig. 3.1. As shown in this figure, field data for more than a year were acquired for each ADN. It should be emphasized that only a very small numbers of studies have been presented in the literature with field data covering a time span of a year or more. However, as already discussed in [12, 25, 26, 39, 40], load/ADN model parameters may significantly vary depending on date and time. Therefore, measuring over at least a year is very important for building accurate and reliable load/ADN models using the measurement-based approach.

During those periods of time, around 1400 measurements on average were recorded at each ADN. For each substation, the measurements are spread out significantly across different months and times of the day in order to include temporal variations in dynamics.

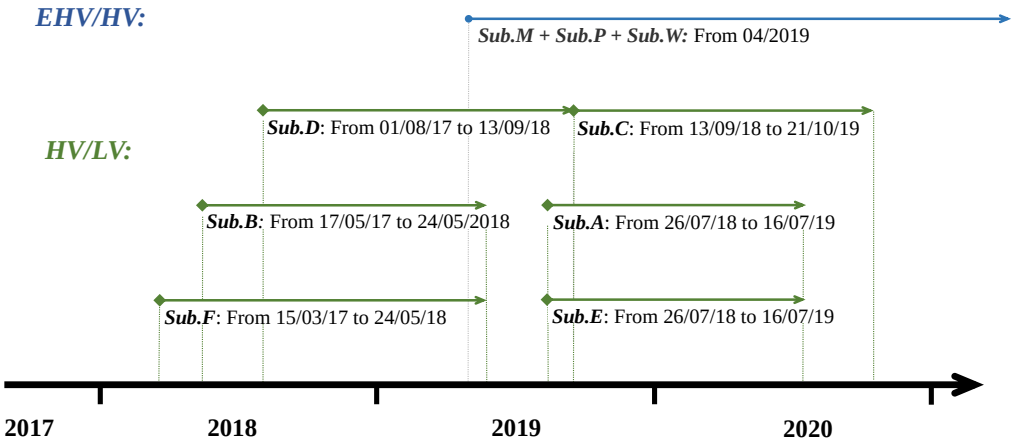


Figure 3.1: Timeline of the start and end dates of measurement data.

3.3 Data preprocessing

3.3.1 Field data

For each ADN of the HV/MV substation list, a dataset $\mathcal{D}$ of M measurements is recorded at the point of common coupling (PCC) based on the trigger criteria described in the "Voltage events" section earlier in this chapter. Each measurement $i = \{1, ..., M\}$ comprises three equally sized vectors $\boldsymbol{V}_i$, $\boldsymbol{P}_i$, and $\boldsymbol{Q}_i$ corresponding to the RMS values of the time domain (TD) signals of voltage, active, and reactive power, respectively. As mentioned, this work examines the dynamic behavior of an ADN under electromechanical transients. During those transients, the bus voltage is expected to vary by $\pm 10\%$ from the nominal value [43], and thus the acquired measurements are expected to contain voltage disturbances within those limits.

A major characteristic differentiating the traditional distribution networks from ADNs is the installed DG. Therefore, the measured power values at the PCC may not always reflect the real load demand. Instead, they express the difference between the real power demand and the power generated by DG. However, accessing detailed DG data in order to estimate the real load demand may be difficult, since there are countries where DSOs do not share those data with TSOs due to confidentiality rules [11]. To this end, a preprocessing phase is performed, in which the initial pre-disturbance conditions $V_{0,i}$, $Y_{0,i}$ are subtracted from the TD values $V_i(t)$, $Y_i(t)$ for voltage and power, respectively, using:

$$\Delta V_i(t) = V_i(t) - V_{0,i} \quad \text{and} \quad \Delta Y_i(t) = Y_i(t) - Y_{0,i}. \tag{3.1}$$

The variable Y can describe either the active (P) or reactive power (Q) while $(V_{0,i}, Y_{0,i})$ can be considered any operating point within a few time steps before the disturbance. This step is fundamental for the generalization capability of the proposed method, since it virtually decouples the dynamic response of the system from its initial unknown power conditions and additionally yields a common reference for the cluster analysis, which is introduced in Chapter 5.

It worth mentioning that generation data could be used in order to estimate the real load consumption. However, this requires high resolution DG data that may

not be available even to the corresponding DSOs. As an alternative, interpolated generation values at the moment of the disturbances could be extracted from the quarterly hour time series power data, which are usually available to the DSOs. In this particular study, many of those time series data were given. Those data correspond to the average $15\,\mathrm{min}$ power values of various DG units installed in the examined ADNs. However, data from a few DG units are missing and thus, the real power demand cannot be correctly estimated.

For the rest of this thesis, the data are presented in pu instead of SI units due to confidentiality reasons. Different base values have been used for each substation. Moreover, the signals corresponding to voltage step-downs ($\Delta V < 0$) were multiplied by -1 only for a visually better and compact representation of the underlying dynamics, as these can be assumed to be linear and thus, symmetrical for voltage disturbances within the range considered.

3.3.2 Weather data

One of the main contributions of this thesis is the the analysis of how exogenous variables can affect the system dynamics. An important part of those exogenous variables is the weather data. Therefore, hourly time series data of temperature, solar radiation, wind speed, and humidity were obtained from the German Weather Agency (Deutscher Wetterdienst - DWD) [112] using the closest weather stations to the examined ADNs. Then, those weather data are interpolated over time in order to estimate the weather conditions at the moment of each disturbance.

3.4 Influence of distributed generation on ADN dynamics

In this Section, the influence of the various DG technologies on the system's dynamic behavior after voltage changes is examined. Importantly, ADNs containing solar and wind parks as well as combined heat and power (CHP) and biogas plants

are assessed. To do so, the recorded measurements are split into two subsets for each of the given transformers. The first subset contains only the entries in which a specific DG technology was not operating whereas the second subset contains the measurements corresponding to the times that the DG unit was generating power. Note that the splitting criterion, i.e., DG operates or not, is based on the interpolated values of the quarterly hour time series DG data provided by the respective DSO. As a consequence, this information is to be considered as partially uncertain.

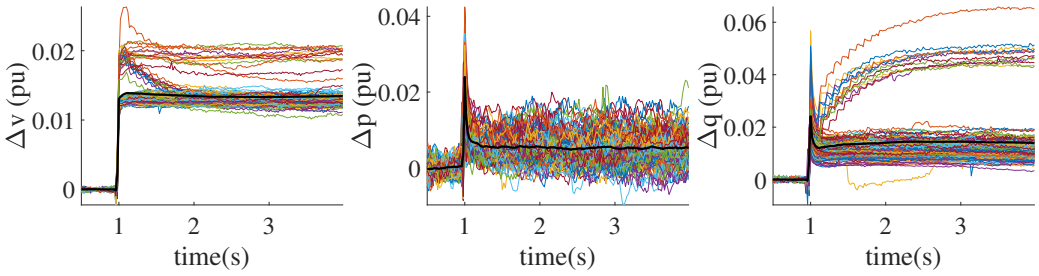
In order to easily compare the influence of the various DG technologies on the dynamics of the system, several plots are introduced for each of the following ADNs. Each ADN is represented by plots of individual measurements of the actual changes in voltage, active, and reactive power as well as their respective ensemble average. In particular, the six-plot figures contain the individual measurements (colored curves) for each derived subset, i.e., one subset with measurements without generation by a specific DG technology (upper plots) and another subset with generation by a specific DG technology (lower plots). For each of those plots, the average curve is shown by a black bold line. To facilitate direct comparison of the subsets, the average curves are also presented in another figure.

3.4.1 Wind parks

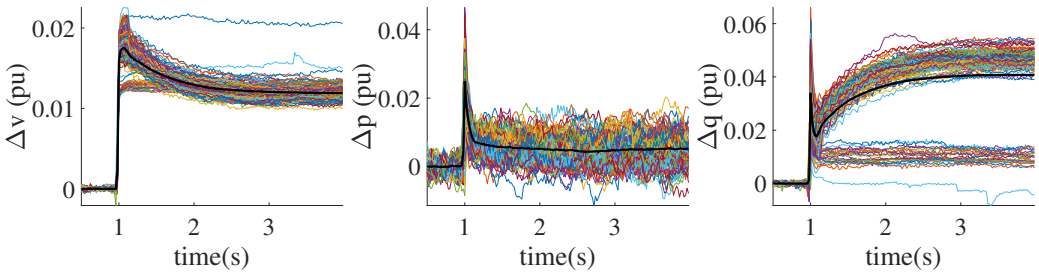
In this section, we examine the influence of wind parks on the system dynamics after a voltage disturbance. To do so, the actual differences of voltage, active and reactive power are assessed using the measurements acquired at Sub.A.1 and Sub.B.1. As mentioned above, the measurements are grouped into two subsets. The first subset contain entries in which the wind parks of the corresponding ADNs were not in operation, whereas the second subset contain measurements in which the wind parks were in operation. The individual measurements of those two ADNs are depicted in Fig. 3.2 and 3.4, while their average curves are illustrated in Fig. 3.3 and 3.5, respectively.

As shown in Fig. 3.3 and 3.5, the average active power responses are very similar in both subsets for both ADNs. The small differences observed may be caused

by the slightly different magnitude of voltage disturbances. At Sub.B.1, for example, the voltage measurements with no wind generation follow a clear step change, whereas the voltage measurements of the second subset are characterized by a quick rise in voltage followed by a small decay until the voltages converges to a new steady state. Regarding the reactive power responses, most of the measurements follow the classical exponential recovery behavior as illustrated in the first subset of both ADNs (Fig. 3.2a and 3.4a). However, the two subsets containing measurements with wind generation (Fig. 3.2b and 3.4b) present different dynamics in comparison with the subsets of no wind generation. This characteristic is clearly visible in the average curves of the reactive power response in the right side of Fig. 3.3 and 3.5. Therefore, both wind parks seem to have a voltage controller which adjusts the reactive power based on the voltage values in a closed control loop. Therefore, the measured voltages are not only influenced by the on-load tap changer position but



(a) Individual measurements acquired with no wind generation.



(b) Individual measurements acquired with wind generation.

Figure 3.2: Sub.A.1: Voltage, active, and reactive power change for the measurements acquired with and without wind generation.

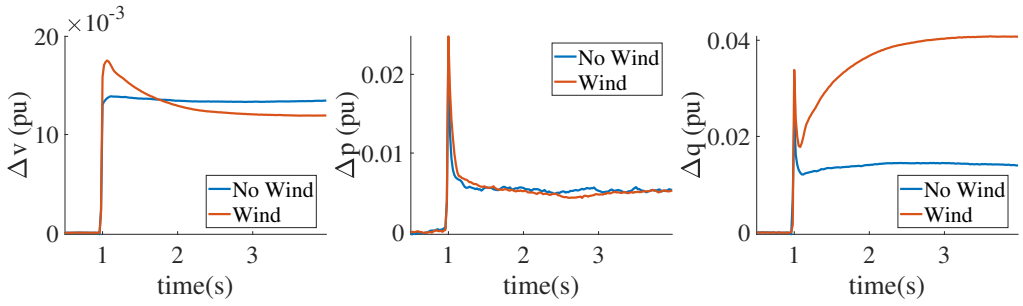
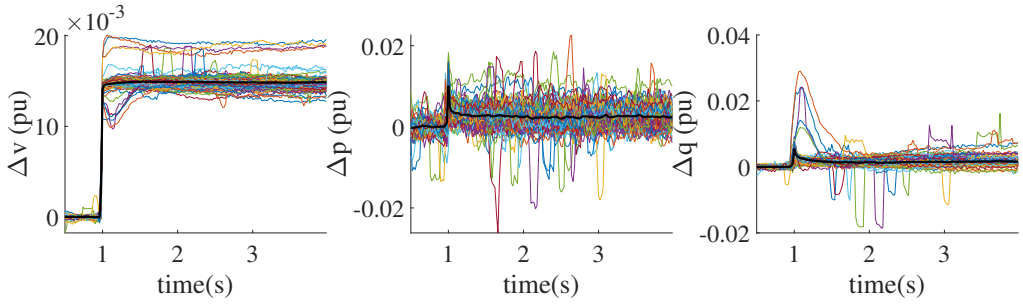
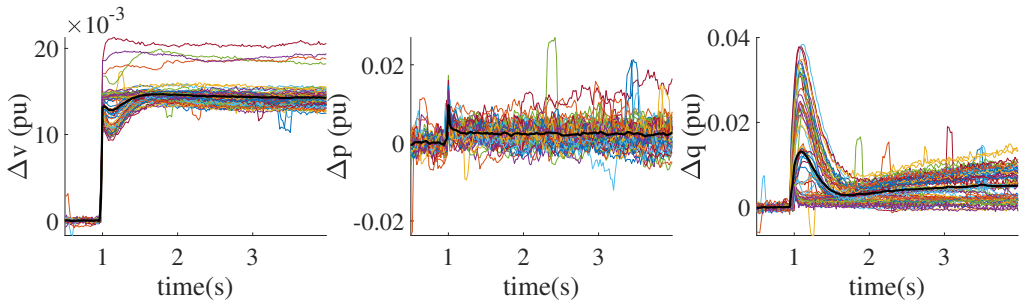


Figure 3.3: Sub.A.1: Average voltage, active, and reactive power change of the measurements acquired with and without wind generation.



(a) Individual measurements acquired with no wind generation.



(b) Individual measurements acquired with wind generation.

Figure 3.4: Sub.B.1: Voltage, active, and reactive power change for the measurements acquired with and without wind generation.

also by the respective voltage controller of the wind parks. It is also worth pointing out that, although both ADNs contain wind parks, those parks seem to be equipped with significantly different reactive power control strategies.

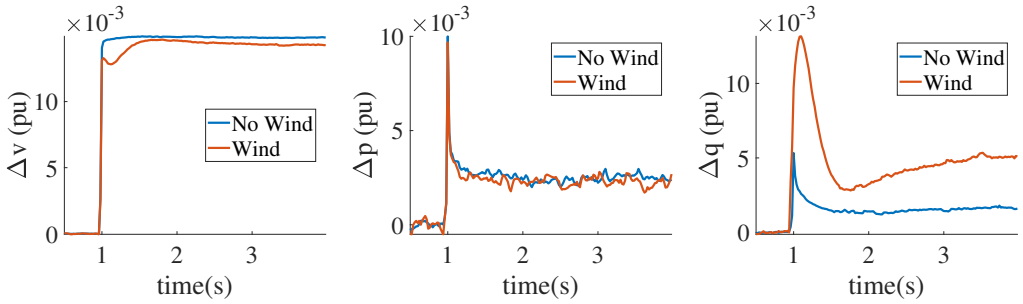
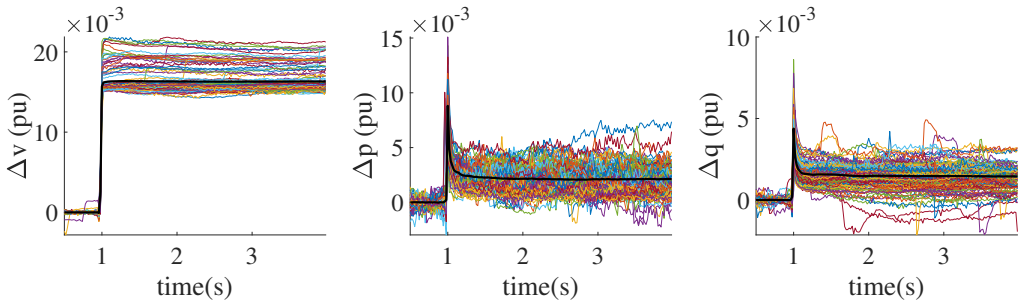
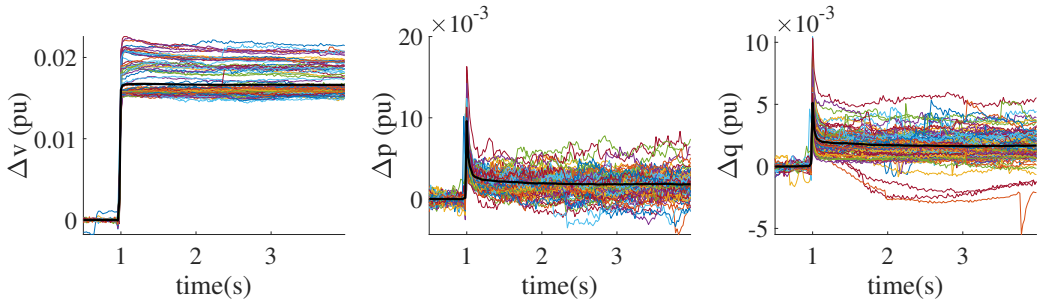


Figure 3.5: Sub.B.1: Average voltage, active, and reactive power change of the measurements acquired with and without wind generation.



(a) Individual measurements acquired with no PV generation.



(b) Individual measurements acquired with PV generation.

Figure 3.6: Sub.A.2: Voltage, active, and reactive power change for the measurements acquired with and without PV generation.

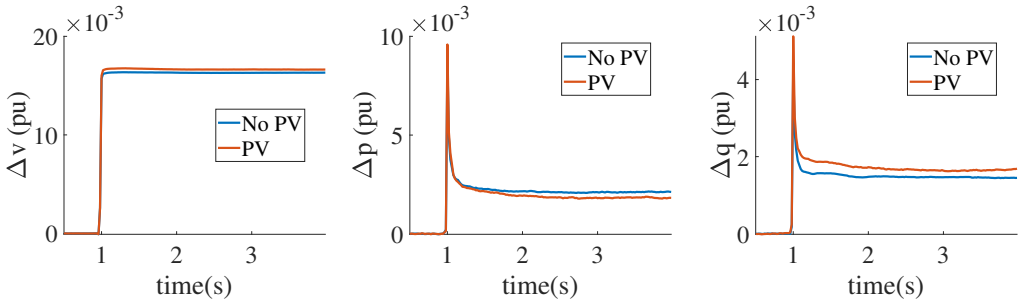


Figure 3.7: Sub.A.2: Average voltage, active, and reactive power change of the measurements acquired with and without PV generation.

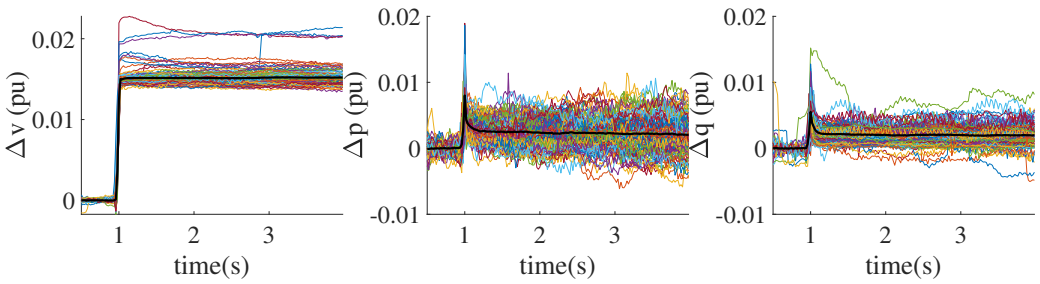
3.4.2 PV parks

In this section, the influence of solar PV parks on the dynamic behavior of the distribution system after a voltage disturbance is assessed. To do so, the measurements acquired at Sub.A.2 and Sub.E.1 are used. The individual measurements of those two ADNs are depicted in Fig. 3.6 and 3.8, while their average curves are illustrated in Fig. 3.7 and 3.9, respectively.

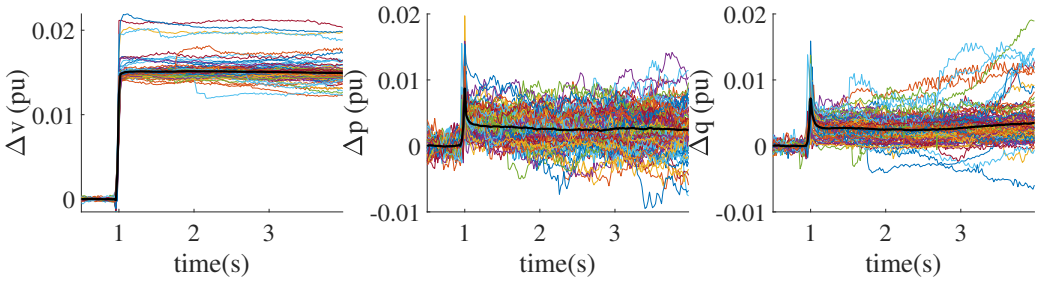
As illustrated in those figures, the average active power responses can be considered as identical in both subsets (No PV, PV), whereas the reactive power responses differ slightly between the two subsets, especially in Sub.E.1. This difference can be partially justified by the highly noisy signals. In general, the PV parks within our examined systems share the same principle in terms of active and reactive power response, meaning there is no influence of PV generation on the dynamic behavior of active and reactive power as observed in the acquired measurements. It should be pointed out here, though, that this conclusion is only valid for the ADNs under study, since other distribution networks may contain PV parks that are equipped with different control strategies.

3.4.3 Biogas plants

In this section, the effects of biogas power plants on the active and reactive power response are evaluated using the measurements at Sub.F.1. The individual mea-



(a) Individual measurements acquired with no PV generation.



(b) Individual measurements acquired with PV generation.

Figure 3.8: Sub.E.1: Voltage, active, and reactive power change for the measurements acquired with and without PV generation.

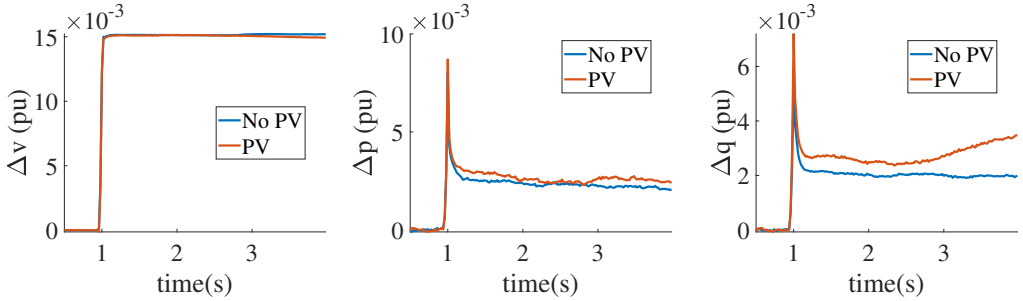
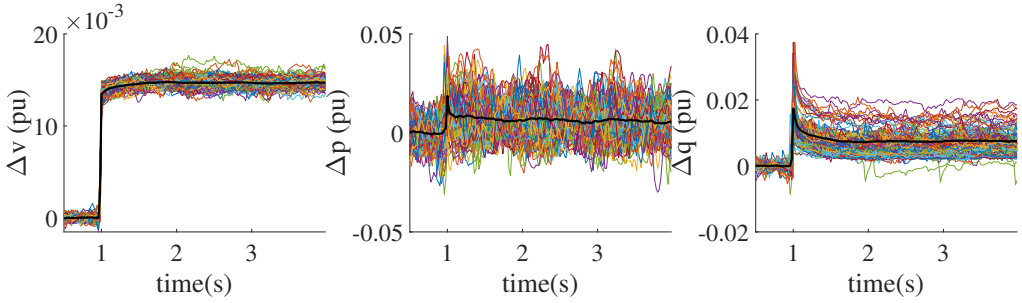


Figure 3.9: Sub.E.1: Average voltage, active, and reactive power change of the measurements acquired with and without PV generation.

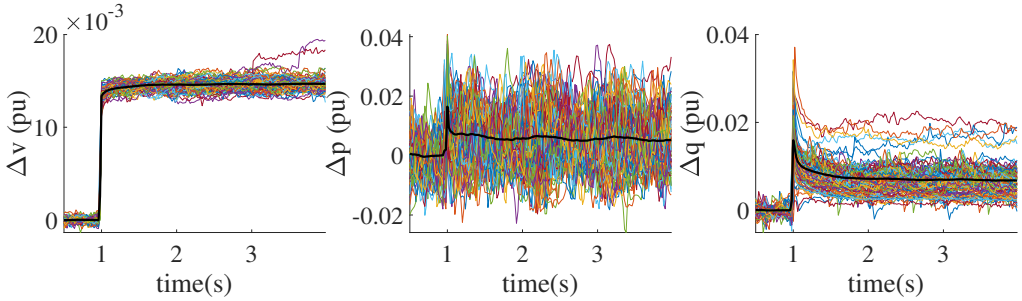
surements of this ADN are depicted in Fig. 3.10, while their average curves are illustrated in Fig. 3.11. As illustrated in Fig. 3.11, the average active and reactive power responses of both subsets are characterized by the same dynamic behavior

and there is no significant difference between them.

It should be mentioned that the measurements of Sub.B.2 follow the same principle, in which the active and reactive power responses are not affected by the biogas power plants. It can be concluded that the active and reactive power of the biogas power plants remain constant after a voltage disturbance. Note that this observation can be attributed to the fact that either those plants do not have voltage support control activated/implemented or they are too small to actually observe it. Importantly, as in case of solar PV parks, this conclusion is valid only in the ADNs under study, since other distribution networks with biogas plants and/or PV parks may include active or reactive power control strategies that influence the dynamics of the system.



(a) Individual measurements acquired with no biogas generation.



(b) Individual measurements acquired with biogas generation.

Figure 3.10: Sub.F.1: Voltage, active, and reactive power change for the measurements acquired with and without biogas generation.

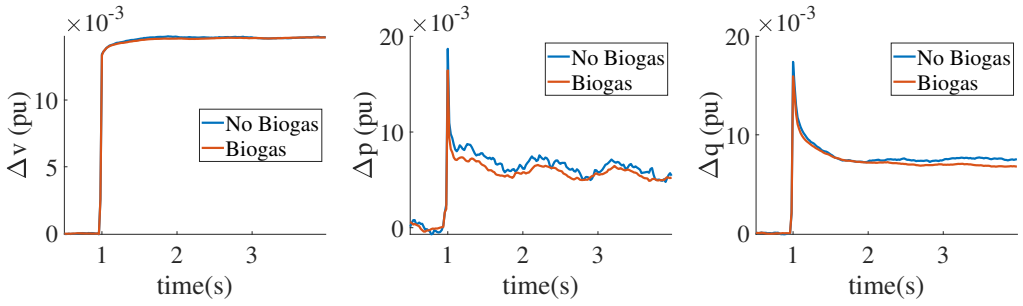
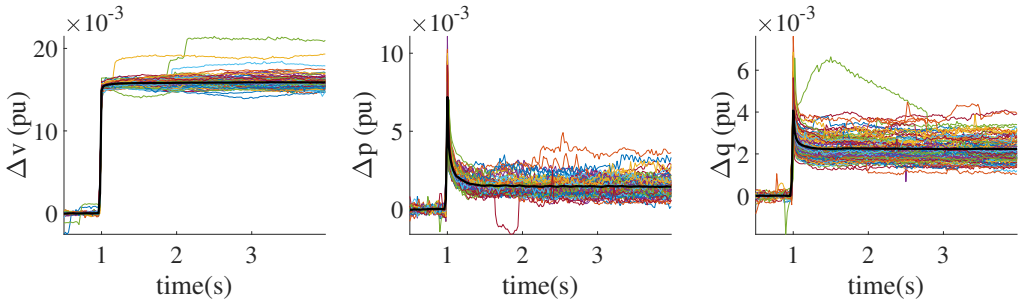


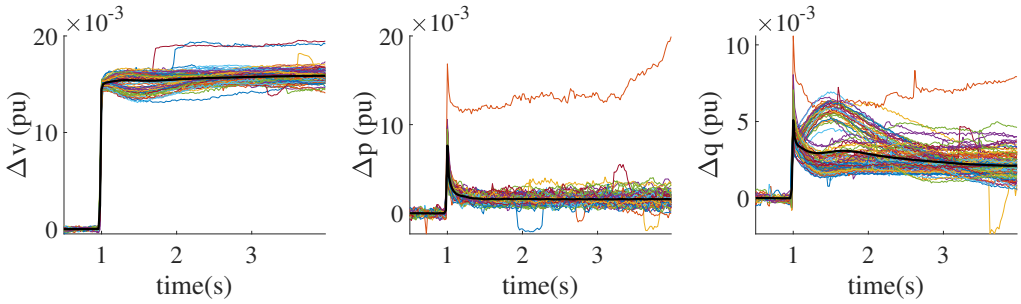
Figure 3.11: Sub.F.1: Average voltage, active, and reactive power change of the measurements acquired with and without biogas generation.

3.4.4 CHP plants

In this section, the effects of CHP plants on the active and reactive power response are evaluated using the measurements at Sub.C.1. The individual measurements of this ADN are depicted in Fig. 3.12, while their average curves are illustrated in Fig. 3.13. In the context of CHP plants, the plots of the two subsets have similar characteristics as the ADNs with wind parks. In particular, the average active power responses of both subsets follow identical dynamic behavior as clearly depicted in the middle plot of Fig. 3.13. On the other hand, the reactive power responses of the two subsets show different dynamics. In the subset with no CHP generation, the reactive power responses follow the classical recovery behavior, in which the power gradually recovers to a new steady-state value. In the subset with CHP, on the contrary, there is a significant number of measurements that show slower dynamics. Furthermore, similar results were observed in case of Sub.D.1, where CHP plants are also present. Therefore, CHP plants seem to have a reactive control concept which adjusts the reactive power based on the voltage values. Again, this conclusion is only valid for the plants considered and other CHP plants may operate without or with a different voltage controller.



(a) Individual measurements acquired with no CHP generation.



(b) Individual measurements acquired with CHP generation.

Figure 3.12: Sub.C.1: Voltage, active, and reactive power change for the measurements acquired with and without CHP generation.

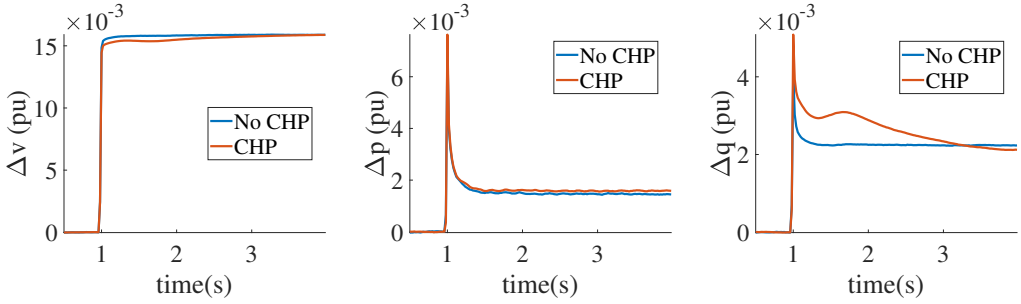


Figure 3.13: Sub.C.1: Average voltage, active, and reactive power change of the measurements acquired with and without CHP generation.

3.4.5 Conclusion

To sum up, no direct influence of PV parks and biogas plants on the dynamic behavior of active and reactive power can be observed in the measurement data, whereas

wind parks and CHP plants seem to have a significant influence on the dynamic responses of reactive power only. In other words, the active power showed no dependence on the DG activity, whereas reactive power responses were influenced by wind parks and CHP plants. However, these conclusions are only valid for the examined substations and cannot be generalized to other ADNs. Depending on the connection code that applied at the time of commissioning, there may be PV parks and biogas plants that perform voltage control, while other wind parks and CHP plants may not.

3.5 Influence of exogenous variables on ADN dynamics

In this section, an exploratory data analysis is performed using two variables extracted from the dynamic load/ADN response and a list of exogenous variables. For the former, we estimate the transient response ΔY_{trans} as well as the steady-state response ΔY_{steady} as

$$\Delta Y_{\text{trans}} = Y_{\text{trans}} - Y_0 \quad \text{and} \quad \Delta Y_{\text{steady}} = Y_{\text{steady}} - Y_0, \tag{3.2}$$

where Y_0, Y_{trans}, and Y_{steady} are the power values before, at the moment, and after the voltage disturbance, respectively, as also depicted in the qualitative plot of Fig. 3.14. For the latter, we employ the temporal and weather features listed in Table 3.4. The temporal features are selected based on the work of [25, 26, 39], which underlines their influence on load dynamics, while the respective weather features correspond to four major factors affecting the DG [113, 114].

3.5.1 Exploratory data analysis

As a first step, it is important to determine how the various measurements are distributed during the day and over the year. In this regard, the measurements should be spread out significantly across different months and times of the day in order to include temporal variations in dynamics. An indicative example of how the recorded

Table 3.4: Exogenous variable list and notation.

Temporal features		**Weather features**	
h	time of the day (hour)	t	temperature
d	day (weekday or weekend)	r	solar radiation
m	month of the year	w	wind speed
		u	humidity

data points are distributed over the day (hour) and the year is illustrated in the violin plot[1] of Fig. 3.15 for Sub.A.1. Although the total number of measurements of this ADN is less than in almost any other ADN of this study, the recorded voltage disturbances are mostly uniformly distributed with the small exception of late night - early morning hours (00:00 - 05:00), where the total number of measurements is smaller than for the rest of the day. Nevertheless, this does not affect the accuracy

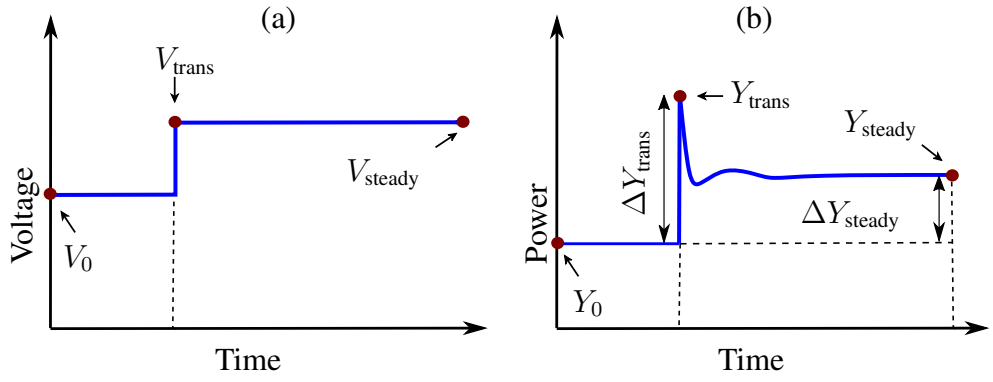


Figure 3.14: Qualitative active or reactive power response of an ADN after a step voltage change: a) Possible voltage step V; b) Indicative (active or reactive) power response Y.

[1] A violin plot shows the probability density of the data at different values as approximated by a kernel density estimator. Similar to box plots, the edge points of the bold black line in the middle of each probability density function denote the 25th and 75th percentiles, while the edge points of the thin black line denote the 0th and the 100th percentiles. The white dot indicates the respective median.

of the developed models, as described in the following chapters. The violin plots of the other ADNs follow the same pattern, where the measurements are uniformly distributed during the day and over the year. Two representative violin plots are illustrated in Fig. 3.16 and 3.17.

At the next step of the exploratory data analysis, we estimate the Pearson correlation coefficients between the aforementioned variables. In general, there is no similar pattern in the correlation values between the dynamics of each substation and the exogenous variables; a result that might be attributed to the different ADN characteristics (load and DG types, transformer capacity). It is worth pointing out that this first step reveals only linear relationships and is performed in order to get an initial understanding of which variables may influence the ADN dynamics.

For instance, the day of the week (weekend or not) has a noticeable influence on the ADN dynamics of Sub.A.1, in particular in case of reactive power, as revealed in Fig. 3.18. For the same ADN, temperature is weakly correlated with the dynamics of both active and reactive power, while humidity seems to have a low impact on the

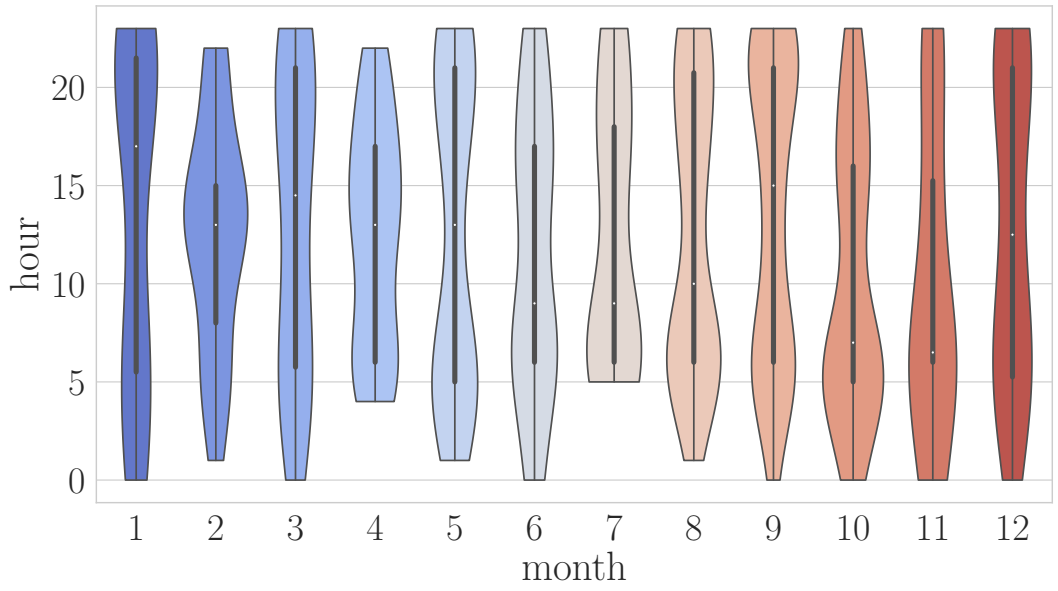


Figure 3.15: Sub.A.1: Violin plot of the total number of recorded voltage disturbances during the day and over the year.

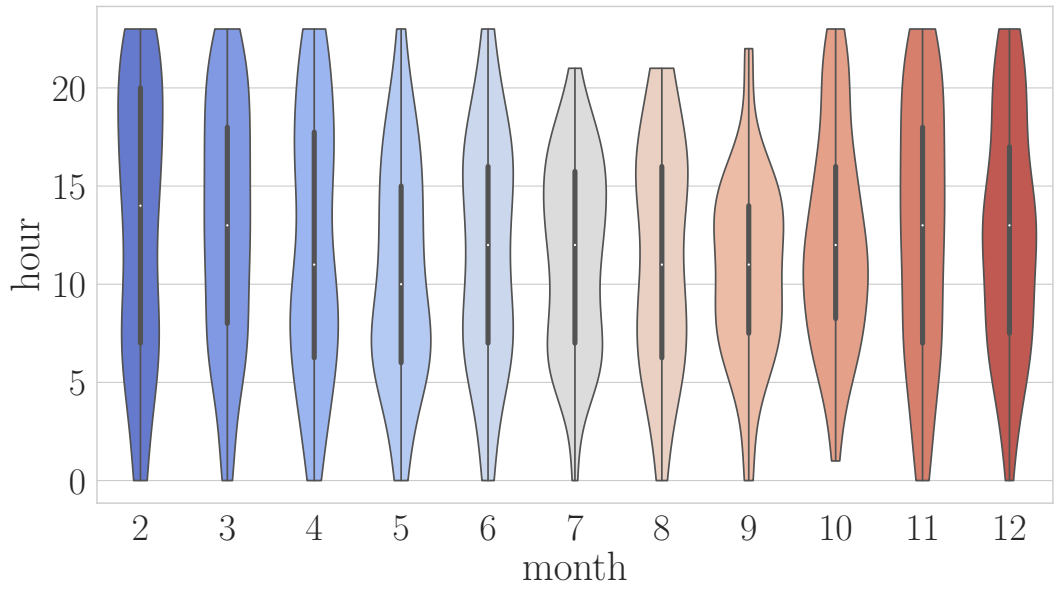


Figure 3.16: Sub.B.2: Violin plot of the total number of recorded voltage disturbances during the day and over the year.

active power responses. In Sub.B.2 (Fig. 3.19), the dynamics of active and reactive power show a significant dependence on the day of the week while the reactive power steady-state response is negatively correlated with temperature. Interesting are the correlations in Sub.C.1 and Sub.D.2, which are presented in Fig. 3.20 and 3.21, respectively. In both ADNs, the active power seems to not be influenced (at least not linearly) by the temporal variables but rather by solar radiation (Sub.C.1) and temperature (Sub.D.2), respectively. Regarding the reactive power of those two ADNs, both of them reveal small-medium dependence on weekdays, temperature, humidity, and radiation, whereas they differ in the influence of month and time of the day (hour). In Sub.C.1, the month of the year affects the reactive power transient response, while the time of the day (hour) seems to have a minor effect on reactive power dynamics. In Sub.D.2, the month of the year appears to have no influence on reactive power whereas the time of the day (hour) show a small yet remarkable linear trend. Finally, one of the biggest linear correlations is revealed between the reactive power dynamics of Sub.E.1 and the day of the week (weekend or not), as

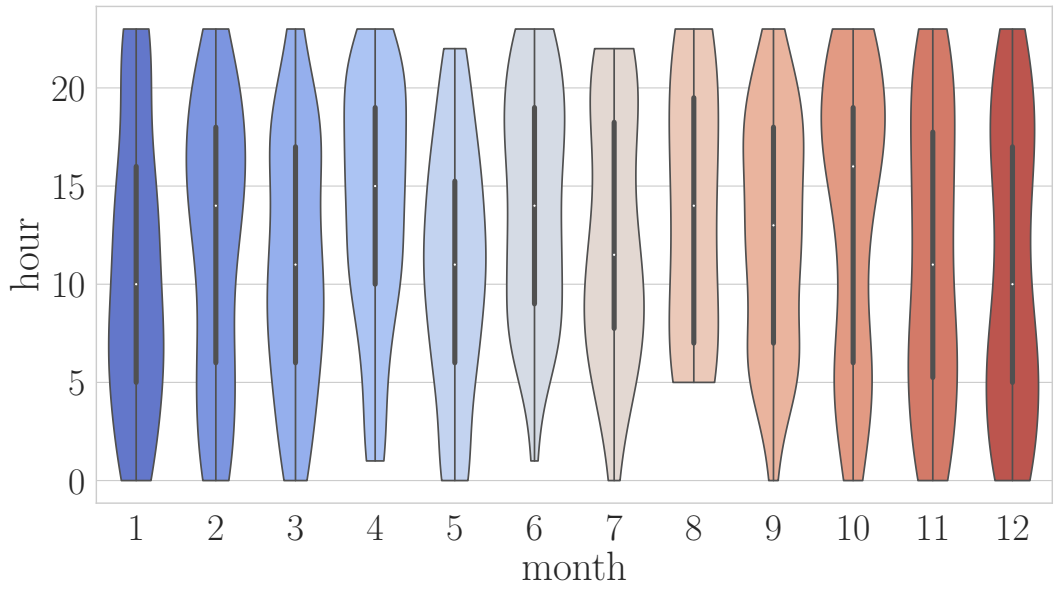


Figure 3.17: Sub.E.1: Violin plot of the total number of recorded voltage disturbances during the day and over the year.

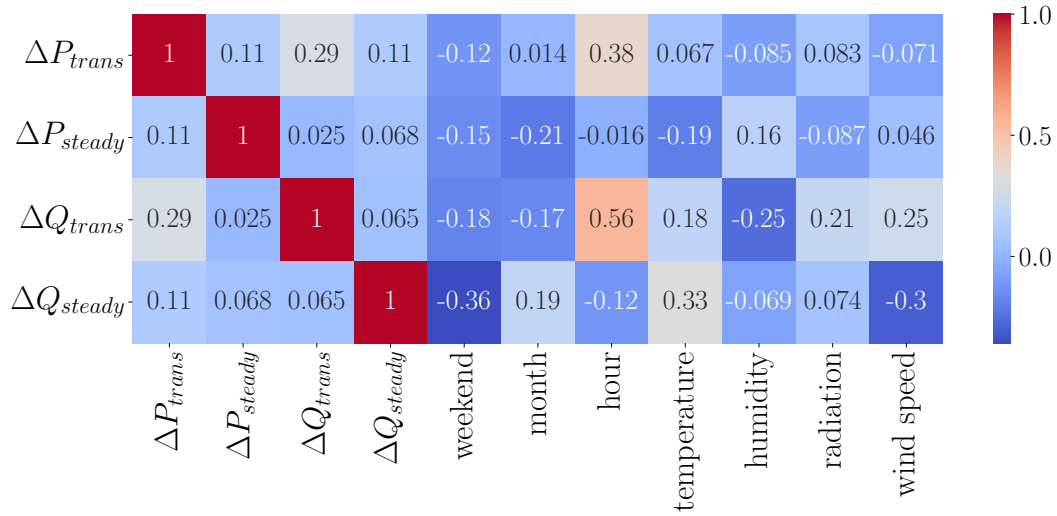


Figure 3.18: Sub.A.1: Correlation coefficients between the various variables describing the dynamics, the time, and weather conditions.

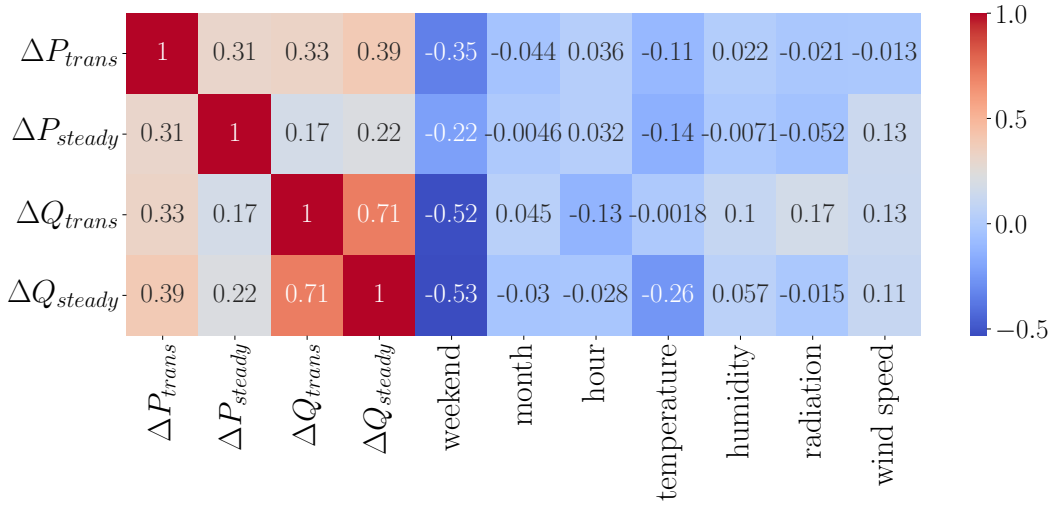


Figure 3.19: Sub.B.2: Correlation coefficients between the various variables describing the dynamics, the time, and weather conditions.

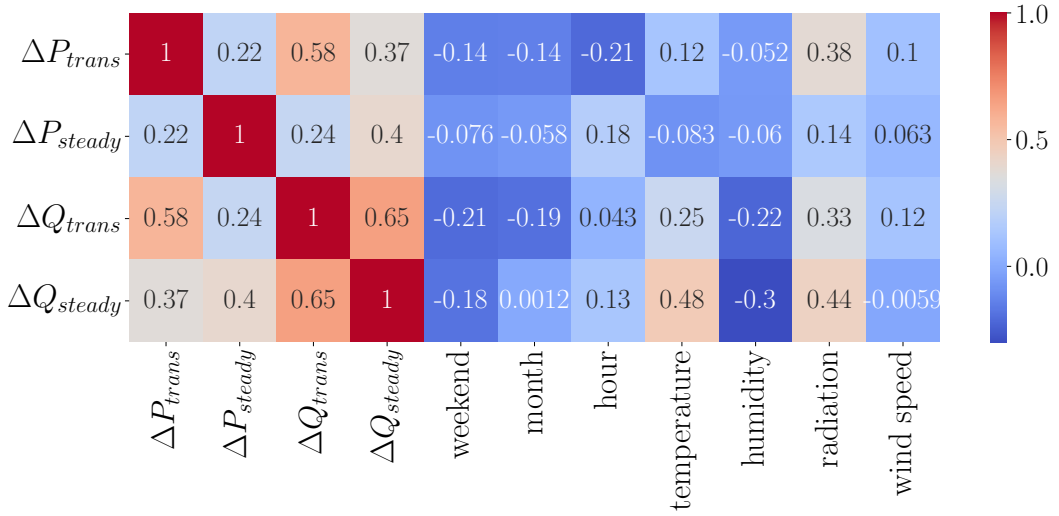


Figure 3.20: Sub.C.1: Correlation coefficients between the various variables describing the dynamics, the time, and weather conditions.

highlighted by the deep blue color in Fig. 3.22. Apart from the day of the week, temperature shows small linear correlation with both active and reactive power.

Nevertheless, the estimated correlation values measure the linear relationships

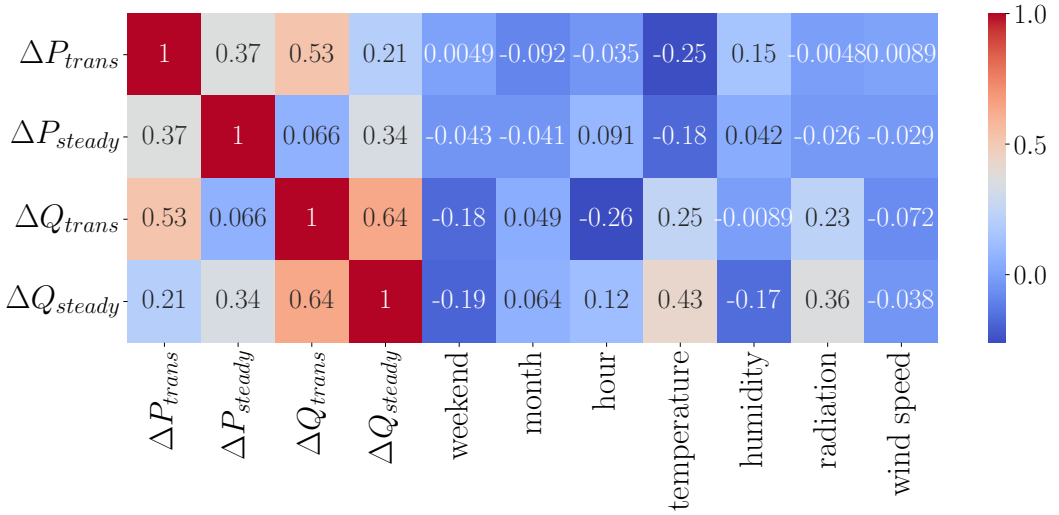


Figure 3.21: Sub.D.2: Correlation coefficients between the various variables describing the dynamics, the time, and weather conditions.

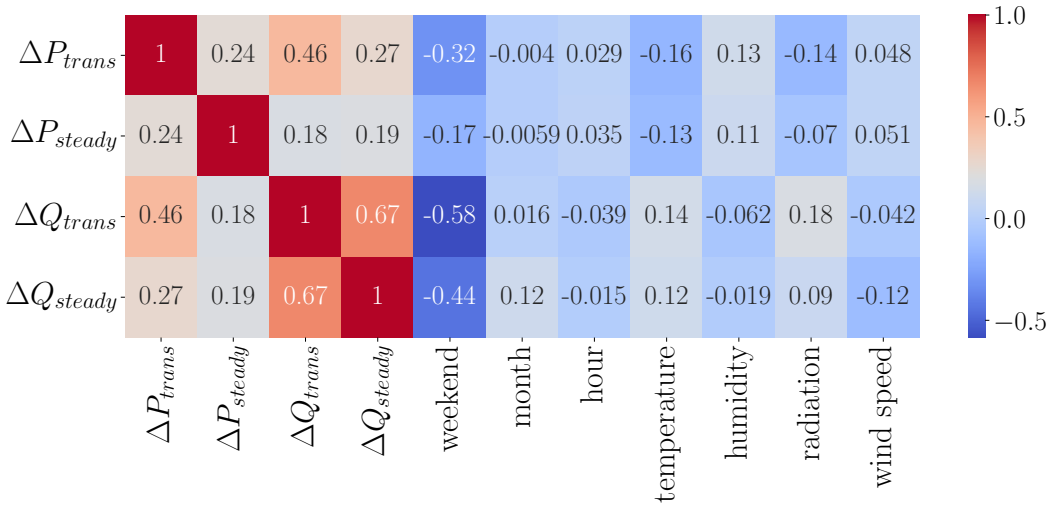


Figure 3.22: Sub.E.1: Correlation coefficients between the various variables describing the dynamics, the time, and weather conditions.

between the given variables and thus, a more detailed data analysis is performed in order to identify possible nonlinear relationships. In this context, we plot the boxplots and the scatter plots of the transient and steady-state power values against

the categorical (day, month, hour) and continuous (temperature, humidity, radiation, wind speed) variables, respectively.

Regarding the categorical variables, various types of relationships between the temporal features and the ADN dynamics have been identified. For instance, the transient active power response of Sub.A.1 shows different values throughout the day, as illustrated in Fig. 3.23. On the contrary, the same transient responses are rather similar with respect to the month of the year (Fig. 3.24). Qualitatively similar results are observed in the transient reactive power response, on which the time of the day has an evident influence whereas the month of the year does not, as depicted in Fig. 3.25 and 3.26, respectively. Another indicative example of nonlinear relationships between the time of the day (hour) and the active or reactive power dynamics is shown in Fig. 3.27 for Sub.C.1. Here, there is a significant increase in the transient active power response during noon hours.

Slightly more complex boxplots are also used in the cases that the day of the week (weekend or not) has a clear impact on ADN dynamics. Representative ex-

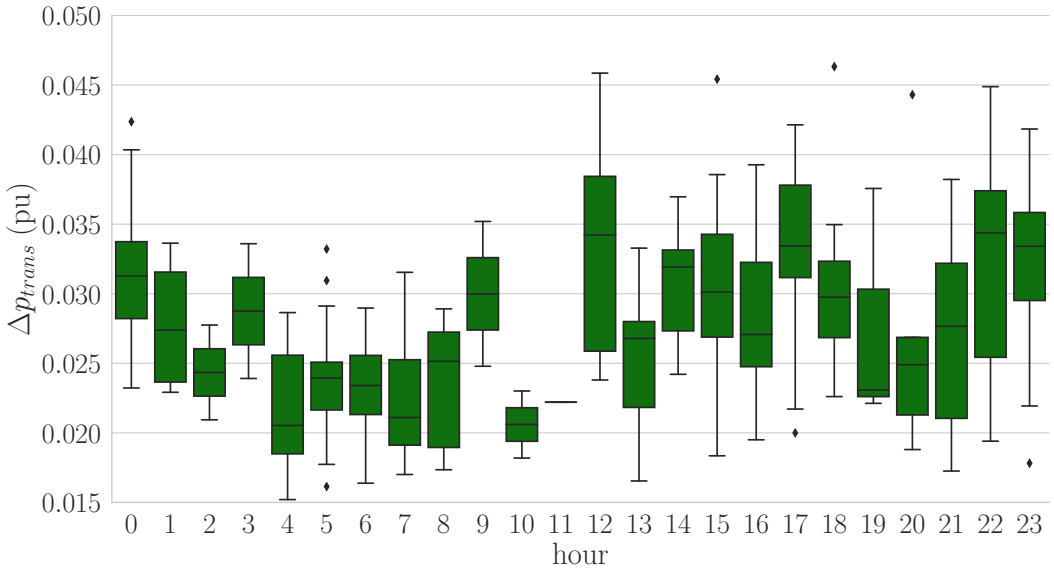


Figure 3.23: Sub.A.1: Boxplots of the transient active power response during the day (hour).

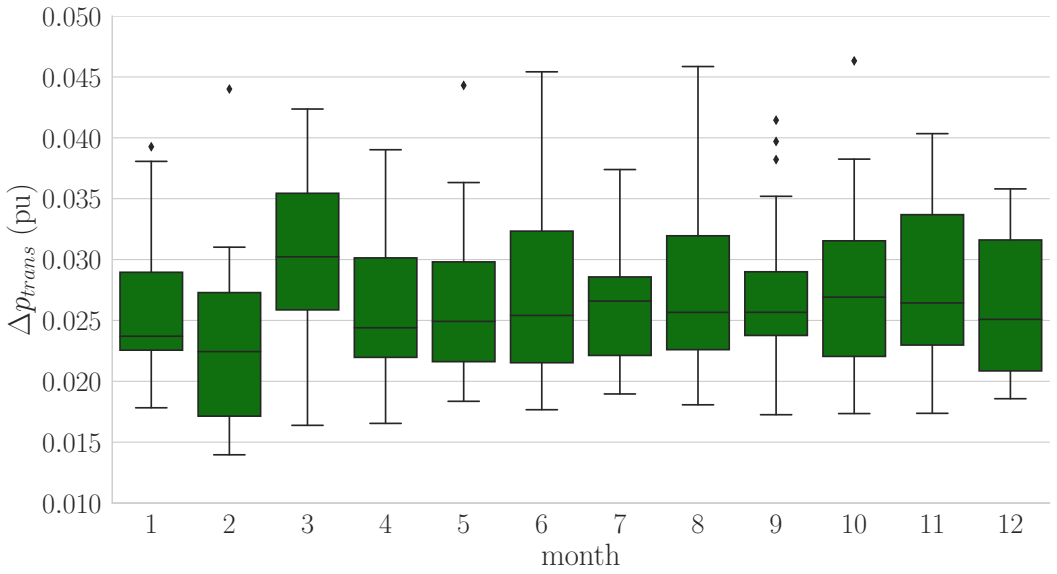


Figure 3.24: Sub.A.1: Boxplots of the transient active power response during the year (month).

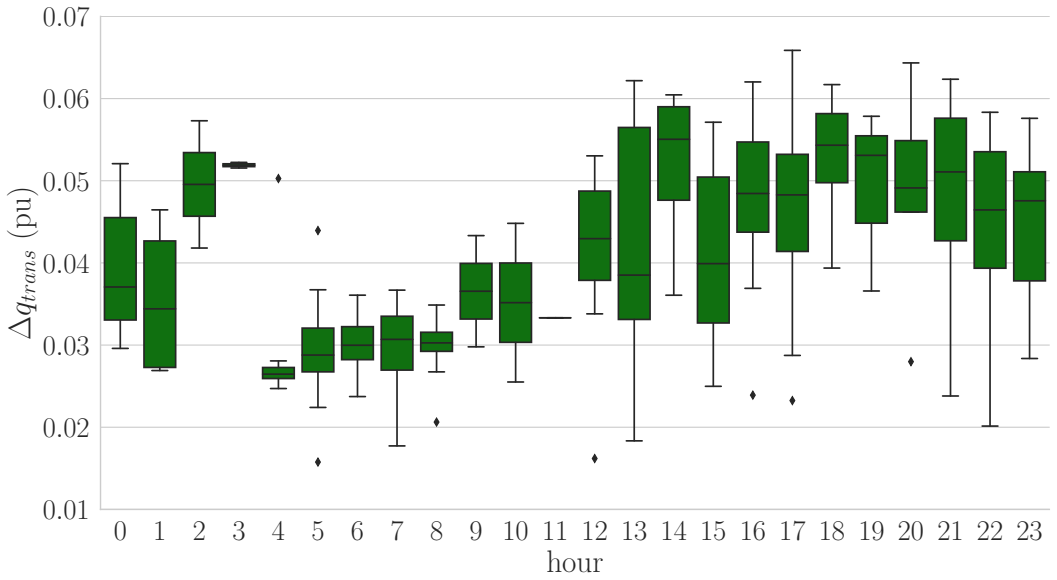


Figure 3.25: Sub.A.1: Boxplots of the transient reactive power response during the day (hour).

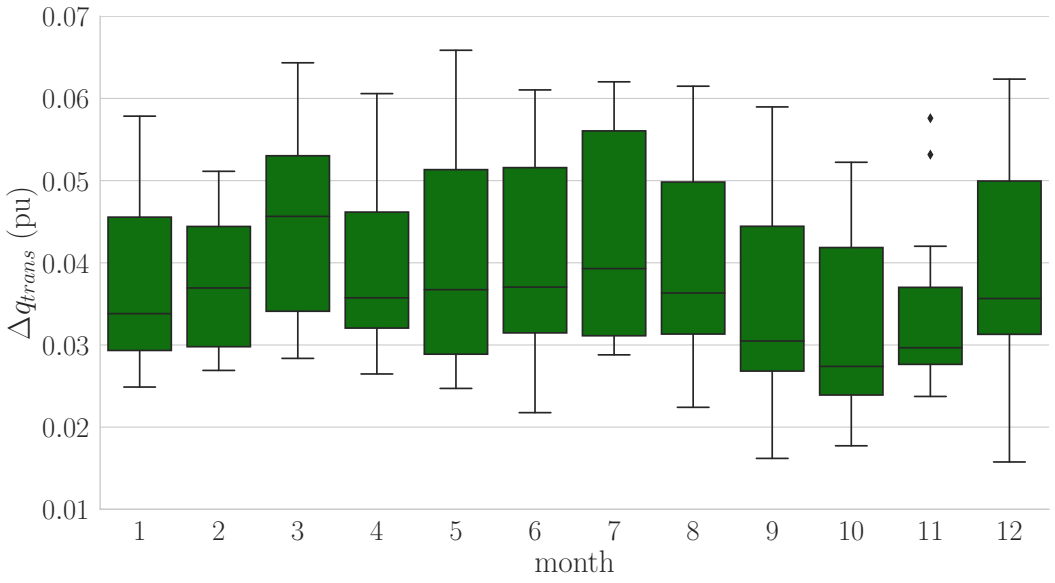


Figure 3.26: Sub.A.1: Boxplots of the transient reactive power response during the year (month).

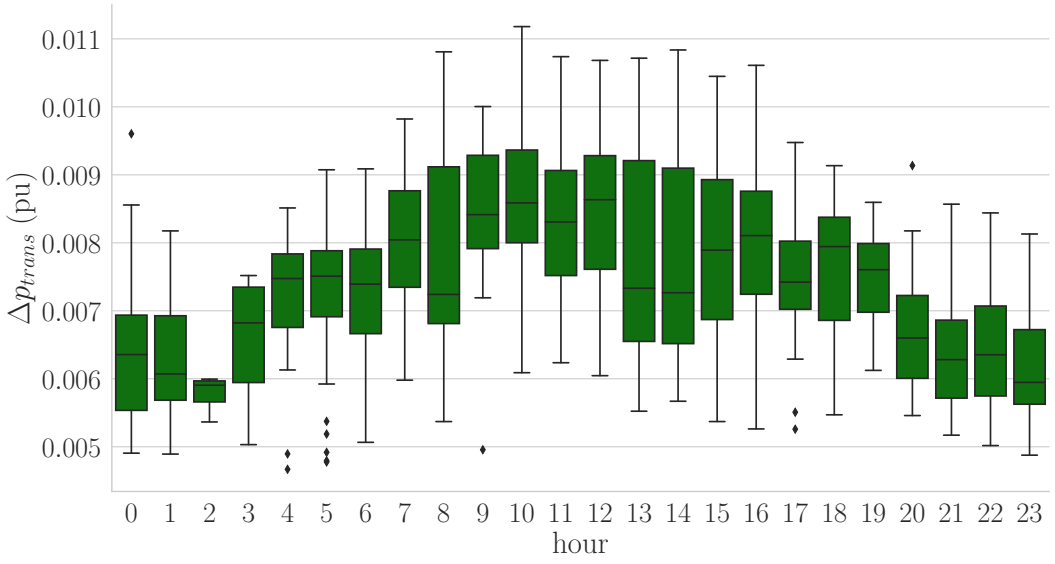


Figure 3.27: Sub.C.1: Boxplots of the transient active power response during the day (hour).

amples of those cases are presented in Fig. 3.28-3.35. In those figures, combinations of the temporal variables manifest an evident impact on the active and reactive power response dynamics. Interestingly, the active and reactive power exhibit a smaller change (deltas) during the weekends compared to the weekdays. This may be attributed to the fact that the total load consumption is usually smaller during the weekends. However, their level of influence varies across the different ADNs under study and no general conclusion can be drawn.

As for the continuous variables, weak linear relationships were predominantly revealed by the scatter plots. Interestingly, small correlations between the reactive power responses and the wind speed appear in a few ADNs with wind parks. This observation may originate from the voltage controllers of the wind turbines, which may operate at a constant power factor. In this scenario, the reactive power will scale with the active power, which naturally depends on the wind speed. An indicative example of this relationship can be seen in Fig. 3.36, where the reactive power change in the steady-state decreases as the wind speed increases.

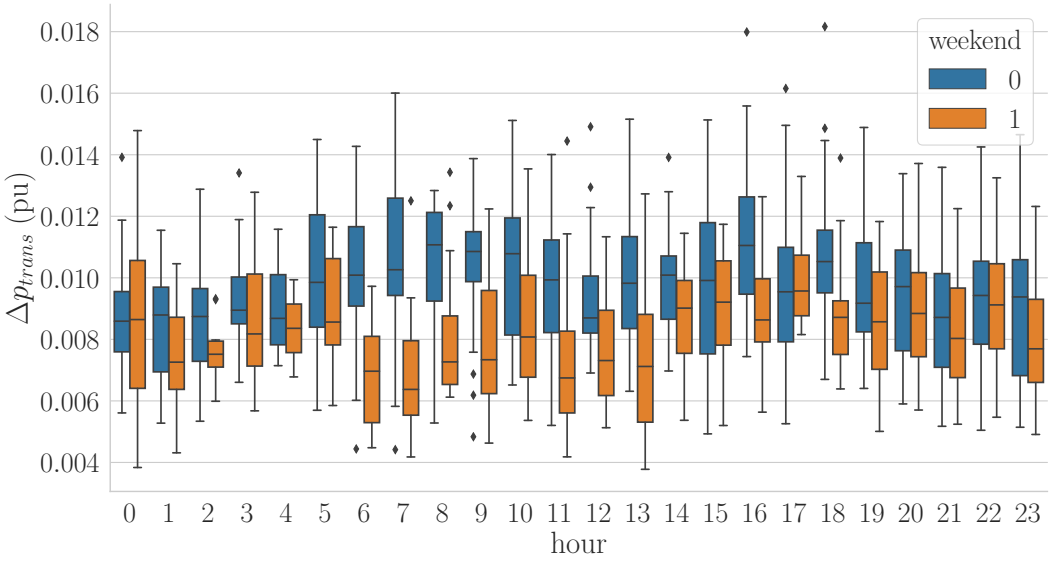


Figure 3.28: Sub.E.1: Boxplots of the transient active power response during the day (hour).

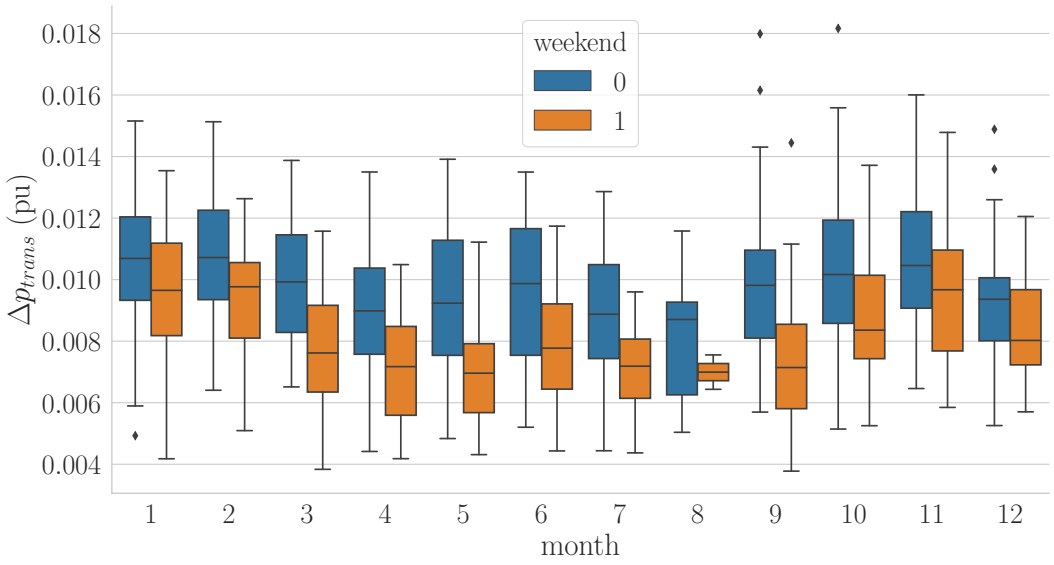


Figure 3.29: Sub.E.1: Boxplots of the transient active power response during the year (month).

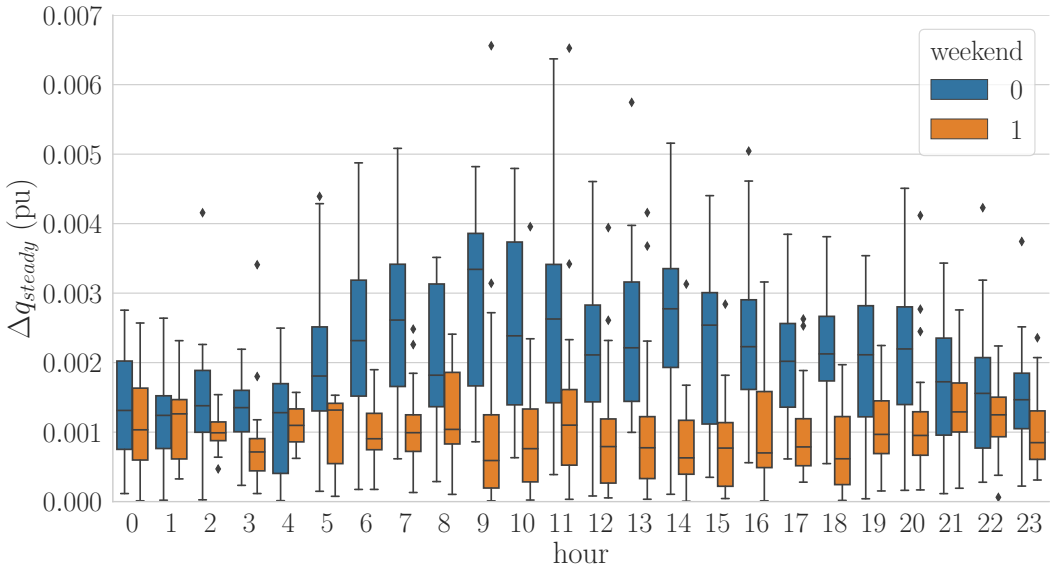


Figure 3.30: Sub.E.1: Boxplots of the steady-state reactive power response during the day (hour).

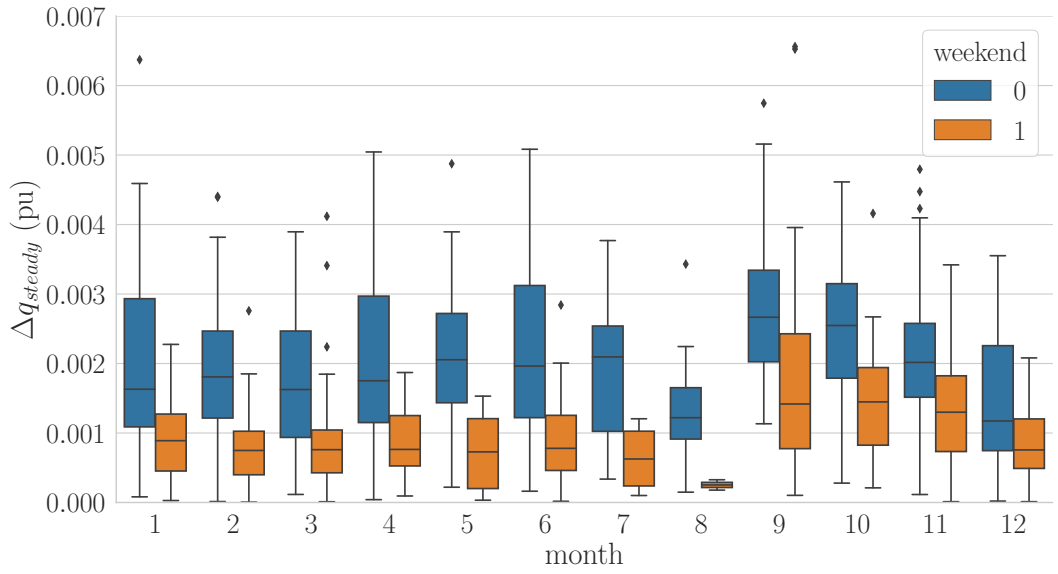


Figure 3.31: Sub.E.1: Boxplots of the steady-state reactive power response during the year (month).

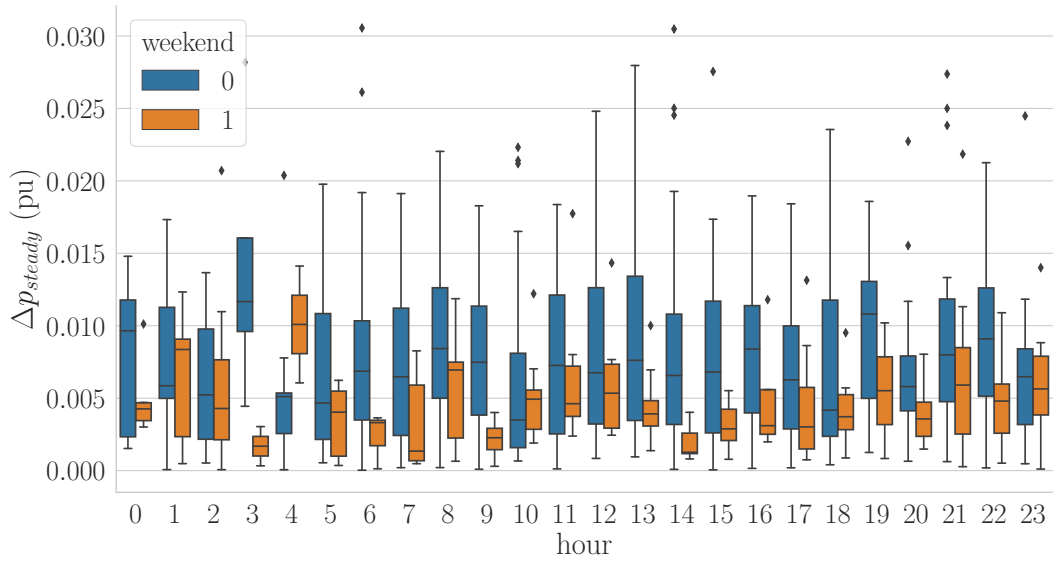


Figure 3.32: Sub.B.2: Boxplots of the steady-state active power response during the day (hour).

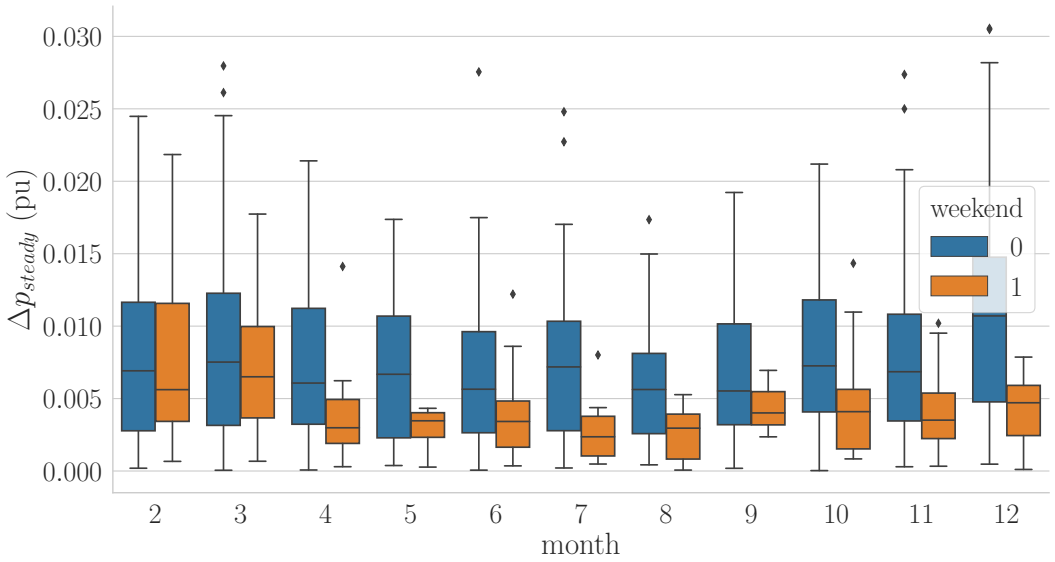


Figure 3.33: Sub.B.2: Boxplots of the steady-state active power response during the year (month).

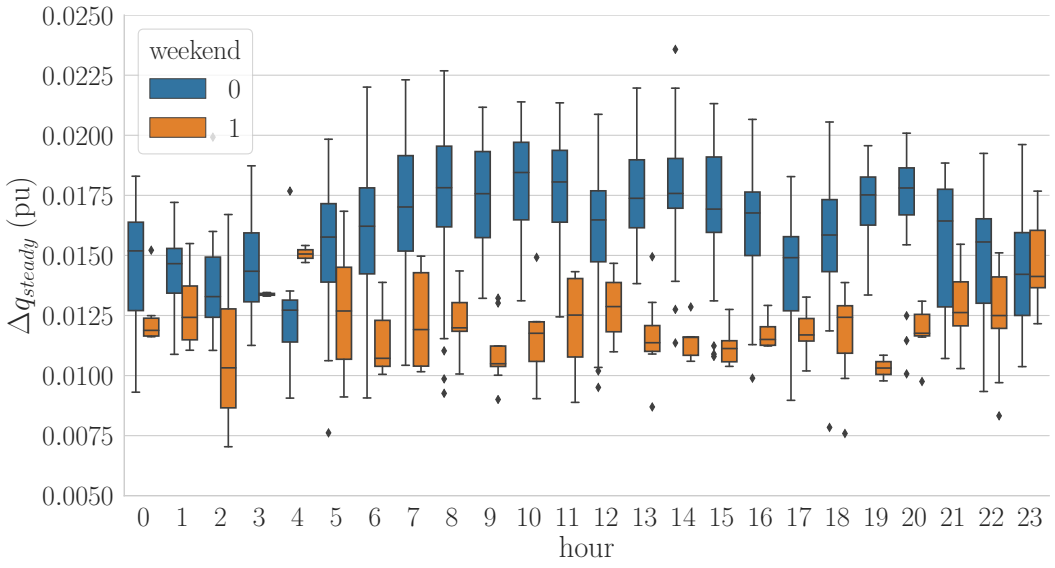


Figure 3.34: Sub.B.2: Boxplots of the steady-state reactive power response during the day (hour).

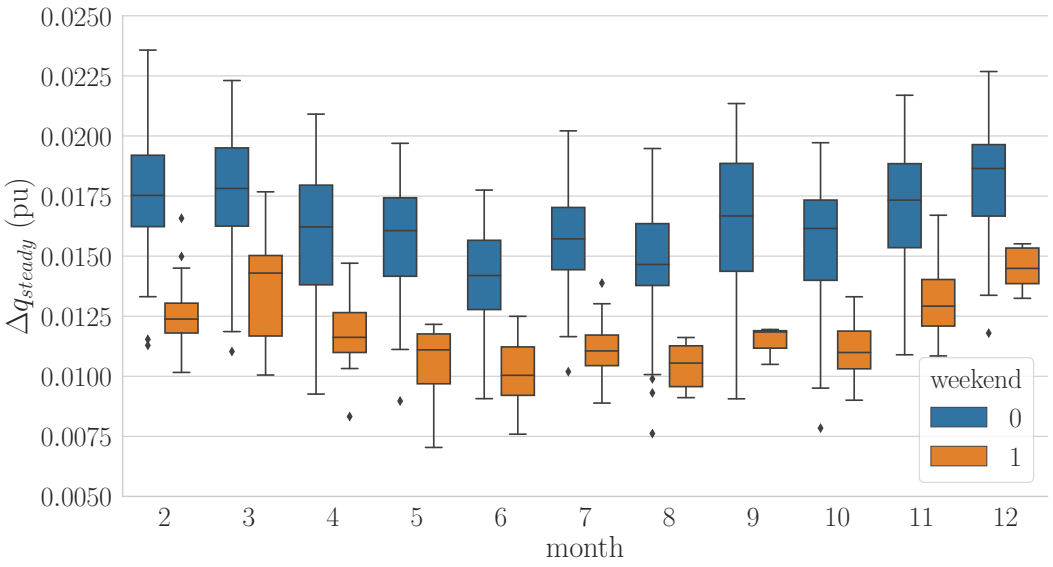


Figure 3.35: Sub.B.2: Boxplots of the steady-state reactive power response during the year (month).

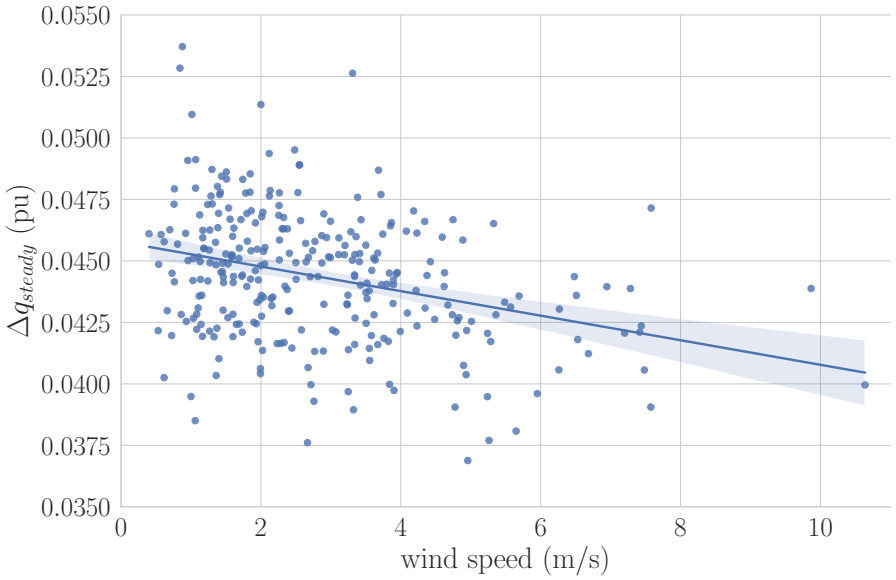


Figure 3.36: Sub.A.1: Scatter plots of the steady-state reactive power response over different wind speed values.

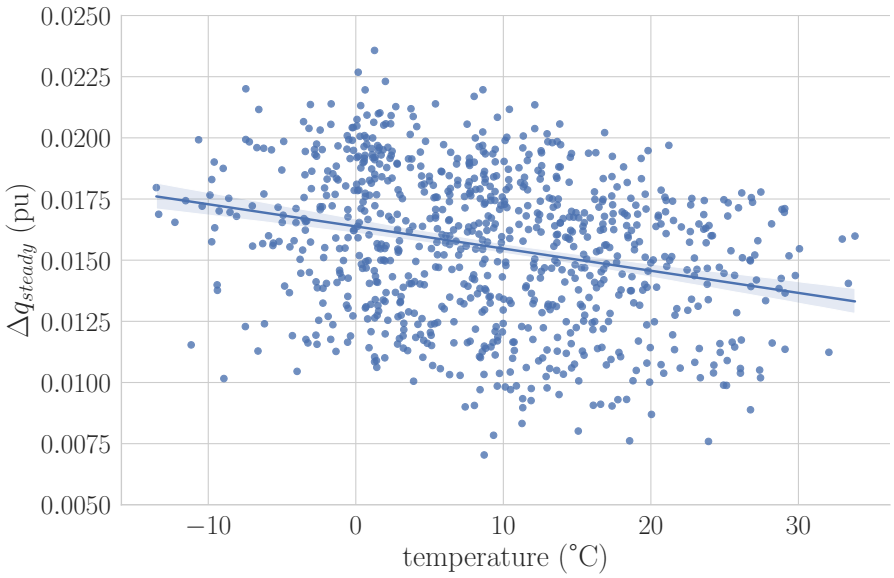


Figure 3.37: Sub.B.2: Scatter plots of the steady-state reactive power response over different temperature values.

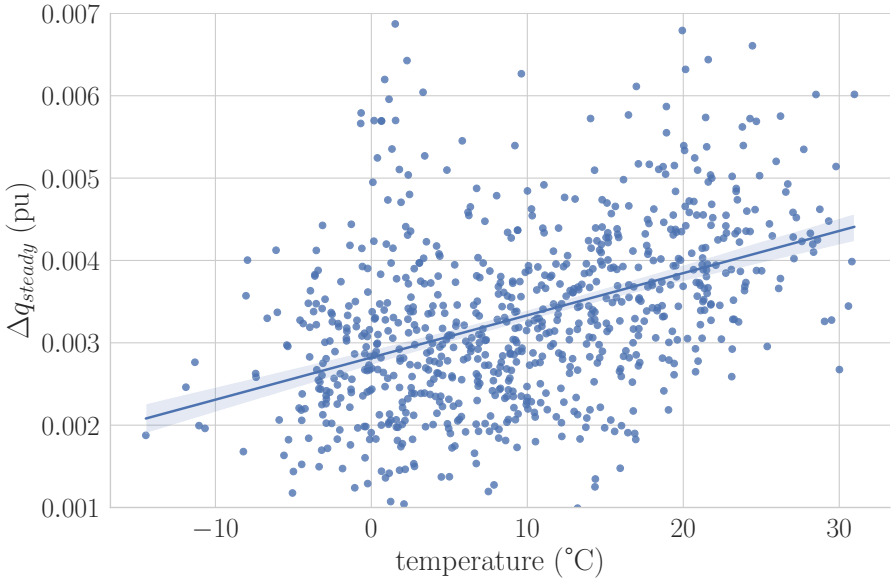


Figure 3.38: Sub.D.2: Scatter plots of the steady-state reactive power response over different temperature values.

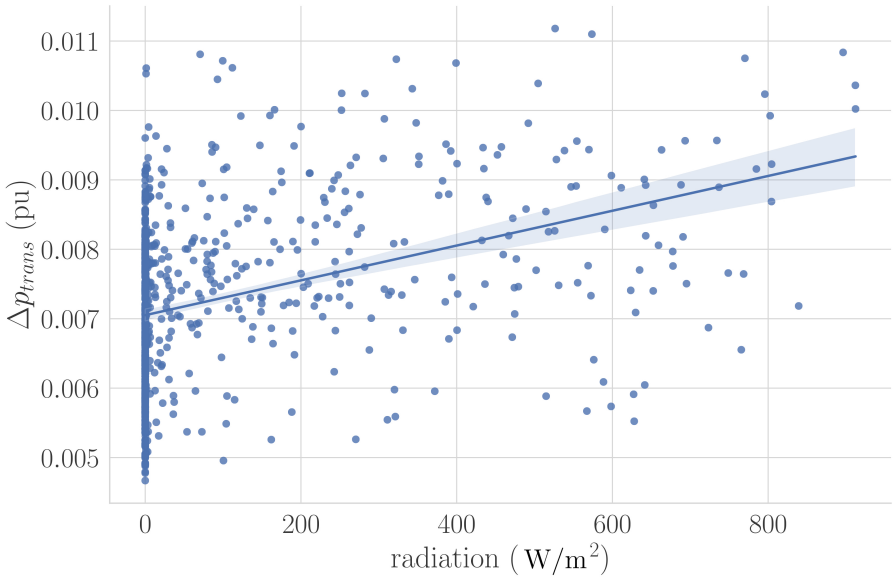


Figure 3.39: Sub.C.1: Scatter plots of the transient active power response over different radiation values.

Furthermore, temperature and radiation seem to have a minor yet worth studying effect on either active or reactive power dynamics whereas humidity does not show a remarkable influence on them. Two scatter plots examples depicting the influence of temperature on reactive power are shown in Fig. 3.37 and 3.38. Sub.B.2 exhibits a negative correlation between the steady-state reactive power response and temperature whereas positive correlation between those two variables is reported in Sub.D.2. As for the impact of radiation, Sub.C.1 shows medium correlation between the transient active power response and the solar radiation, as presented in Fig. 3.39, even though there is not any PV park within this ADN.

3.5.2 Conclusion

The presented analysis is one of the very few studies aiming at revealing relationships between the ADN dynamics and a list of exogenous variables. In particular, the transient and the steady-state response of active and reactive power were employed in order to encode the ADN dynamics in single values. After a thorough

analysis, it was observed that temporal variables, such as time of the day, day of the week (weekend or not), and month, may have a significant influence on the system dynamics. Nevertheless, no general patterns have been identified and each ADN exhibits different temporal correlations, which are highly nonlinear. At the same time, weather variables seem to play a minor yet worth considering role in the ADN dynamics. In this context, several ADNs showed small to medium correlations with temperature, radiation, and wind speed whereas humidity seems to have a very minor influence. However, each ADN unveiled different relationships with respect to the weather variables and thus, no general conclusion can be extracted. That would probably require data from significantly more substations.

Those data analysis results highlight the high complexity of developing generic load/ADN models. Based on the existing literature described in Chapter 2, there is no model structure that allows the incorporation of exogenous variables into the model calculations. Therefore, there is a need for new model structures that can decode the information encrypted in the temporal and weather variables and translate it into more accurate results.

3.6 Summary

In this chapter, the main measurement requirements were identified based on the measurement-based approach. Regarding the voltage events, those requirements include a triggering threshold of 1% voltage change, a measurement period of $10\,\mathrm{s}$, and a sampling frequency of $100\,\mathrm{Hz}$. As for the frequency events, frequency does not experience step changes, even in cases of large active power imbalances. Therefore, continuous measurements with a time resolution of $2\,\mathrm{s}$ were selected. This flexibility in setting the type of measurements, e.g., triggered, continuous, or the triggering criteria was provided by the units *PQI-DA smart* (frequency events) and *PQ-Box 150* (voltage events). Both of them are manufactured by *A. Eberle GmbH & Co. KG*.

Special emphasis was put on selecting the examined ADNs. To increase the

signal-to-noise ratio, the units for the frequency events were installed in three EHV/HV substations in order to measure a large grid ("network group"). In the case of voltage events, six utterly different HV/MV substations were selected. The underlying ADNs differ significantly from each other with respect to the load types and the installed DG technologies. Importantly, each measurement unit was collecting measurements for over a year, resulting in around 1400 measurements on average for each ADN.

Due to the fact that the measurements are acquired at the PCC, the measured power values do not reflect the real power consumption. The measured values actually express the the difference between the real load demand and the power generated by DG. To this end, a preprocessing phase is performed in order to virtually decouple the system dynamics from the unknown initial power values. To do so, the actual difference between the recorded signals and the initial pre-disturbance values are calculated for voltage, active, and reactive power.

In the last part of this chapter, we study how DG and a set of exogenous variables may influence the system dynamics. It was observed that none of the DG technologies have a direct impact of active power dynamics. Similarly, the PV parks and the biogas plants seem to have no influence on reactive power dynamics either. On the contrary, the wind parks and the CHP plants are equipped with controllers that support the voltage through reactive power injections. However, those conclusions are only valid for the examined substations and cannot be generalized to other ADNs. As for the temporal and weather variables, linear and nonlinear relationships have been revealed for each substation. Nevertheless, each ADN exhibits different dependence on weather and time and thus, no general conclusion can be extracted.

Chapter 4

Analysis of Frequency Events

The aim of this chapter is derive the steady-state relationship between the frequency and active power, i.e., *self-regulating effect*, based on the frequency and power measurements from the three EHV/HV substations. During the measurement period, many inadequate measurements were incorrectly detected as a frequency event due to the limitations of the trigger criteria that can be set in the used devices. As a consequence, the related data had to be preprocessed, e.g., data cleaning, outlier detection, before any further analysis could be performed.

Interval data of voltage, frequency, active and reactive power with a sampling period of $2\,\mathrm{s}$ were acquired for all three EHV/HV substations. Based on those data, two different methodologies are tested in order to estimate the steady-state relationship between frequency and active power. While certainly of interest, obtaining an estimate for the dynamic relationship between frequency and active power from the data turned out to be infeasible. This is due to the fact that the measurements do not contain a frequency event in which the frequency changes fast enough with sufficient amplitude. For this reason, it is not possible to separate the change in active power induced by the change of frequency from the noise in the active power signals.

The first methodology is based on the frequency change criteria as introduced in [65]. The second approach focuses on the deterministic frequency deviations, which occur hourly every day. In both methods, the power values of all three substations

are added up in order to calculate the total power consumption of this network group and to improve the signal-to-noise ratio. It should be mentioned that there were data with negative active power measurements which were caused by unknown reasons[1]. Those data were considered as incorrect (as there is insufficient DG within this grid to cause reverse power flows) and were discarded from the following analysis. Furthermore, some of the measurement devices were not available at some times. As a consequence, the graphs presented within this chapter can illustrate either the total power consumption of all three substations or the power of two or even one substation. Nevertheless, it will become clear at the end of this chapter that those issues do not affect the following analysis.

4.1 Frequency change criteria

This method searches through the whole dataset of measurements in order to find frequency changes greater than a user-defined threshold within a time period of $3\,\mathrm{s}$. In particular, frequency gradients greater than the values covered by the light blue region in Fig. 4.1 are stored for further processing. In that case, frequency and active power signals are stored for the interval $t_0 - 60\,\mathrm{s} \leq t \leq t_0 + 120\,\mathrm{s}$, where t_0 is the time of the disturbance. Note that the frequency change is calculated as:

$$\Delta f_i(t) = f_i(t) - f_{0,i}, \tag{4.1}$$

where i is the measurement index, $f_i(t)$ denotes the recorded frequency signal, and $f_{0,i}$ indicates the frequency value at t_0.

The search results yielded the signals shown in Fig. 4.2. The left plot represents the stored frequency changes, whereas the right one denotes the corresponding active power signals. After a detailed examination of each individual active power curve, the following conclusions were derived. There may be significant power

[1] A possible explanation could be incorrect assumptions with respect to transformer switch states, as these had to be reconstructed from switching protocols provided by the TSO and may not be fully reliable.

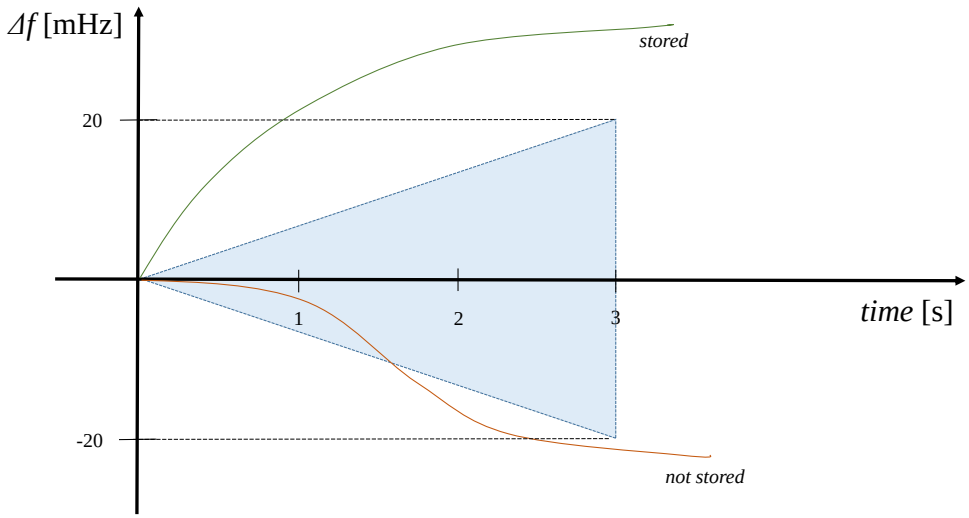


Figure 4.1: Frequency criteria for storing a measured signal.

changes at the time steps of the $15\,\mathrm{min}$ products of the electricity market. Those active power changes are too big to be considered as a self-regulating effect and they correspond to the generation/consumption scheduling as defined by the electricity market. Two indicative examples of this behavior are shown in Fig. 4.3.

In Fig. 4.3a, the total power consumption is around $19.6\,\mathrm{MW}$ and at 20:45, frequency drops by almost $50\,\mathrm{mHz}$. At the same time, there is a $0.6\,\mathrm{MW}$ decrease in active power, which is caused rather by a scheduling of generation units and large consumers than by the frequency drop. Similar results are observed in Fig. 4.3b. Here, a $60\,\mathrm{mHz}$ frequency drop corresponds to a $2\,\mathrm{MW}$ decrease in active power, while the total power consumption is around $105\,\mathrm{MW}$. To conclude, at the time steps of $15\,\mathrm{min}$ products, any self-regulating effect is overlapped by a bigger change in active power due scheduling of consumers and generation units. In fact, the partial scheduling of generation and load within the examined grid contributes to system imbalance and, hence, to the frequency deviation.

Since frequency changes at the time steps of $15\,\mathrm{min}$ products have shown not to be particularly useful when individual measurements are used, the rest of stored sig-

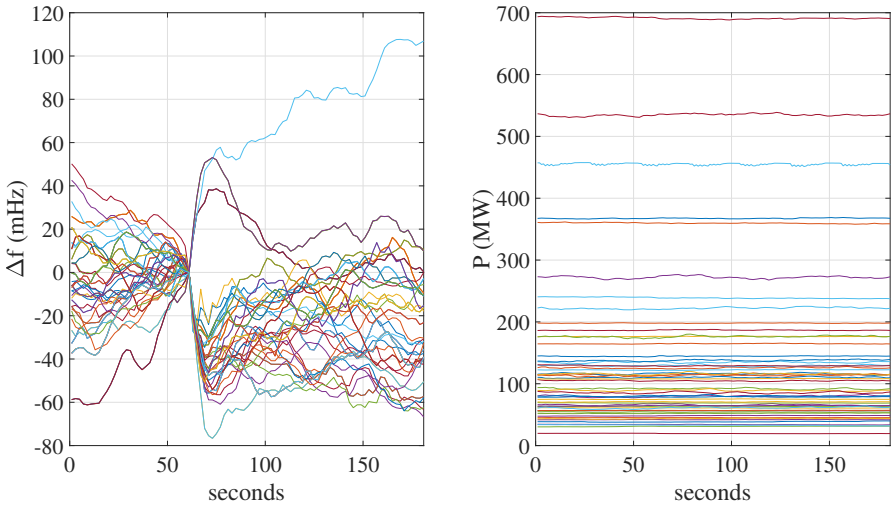
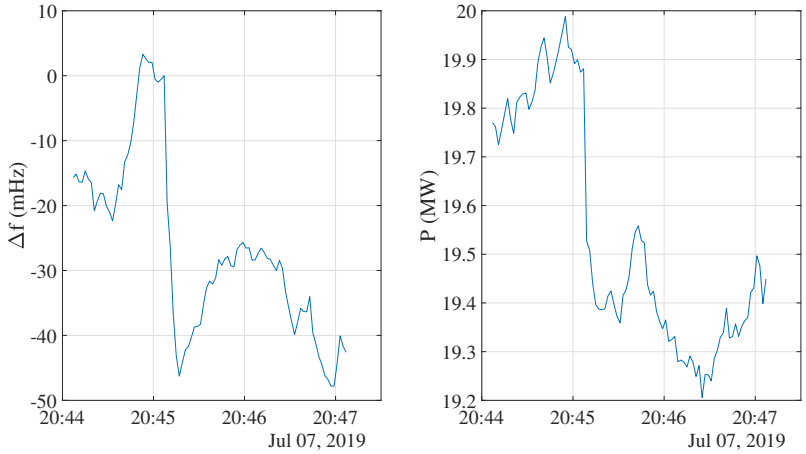


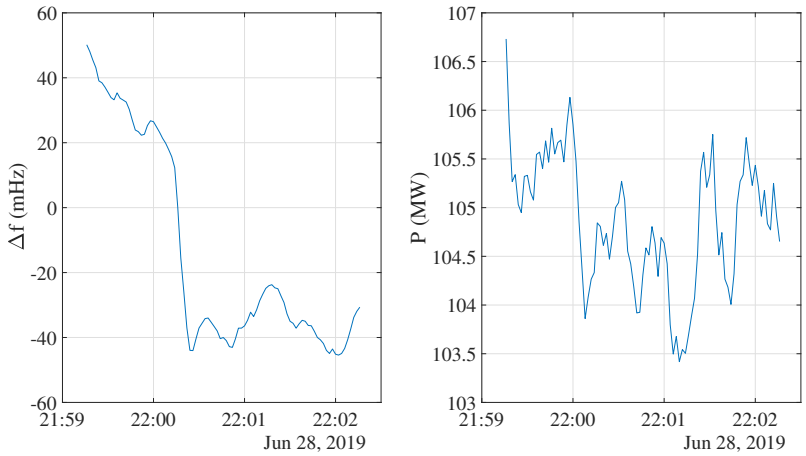
Figure 4.2: Stored frequency and active power signals.

nals is examined. Examples of those measurements are presented in Fig. 4.4-4.11. It is clear that the active power signals contain noise coming from the switching of the consumers and the changes in frequency are not followed by similar changes in active power. Therefore, individual measurements cannot be employed in practice, since the signal-to-noise ratio is relatively low and no meaningful insight into the self-regulating effect can be extracted. To this end, an ensemble averaging of active power signals was performed over two subsets. The first one refers to the signals with positive frequency change, whereas the second subset contains the measurements with a negative frequency change. The motivation for using ensemble averaging is based on the intuition that the noise in active power signals will cancel out partially and thus, the signal-to-noise ratio will increase.

In Fig. 4.12 and 4.13, the black bold curves denote the average signals of frequency and active power for negative and positive frequency changes, respectively. It is clear that there was no significant change in the active power signal caused by the change in frequency and the stochastic part overlaps any self-regulating effect. It can be concluded that the self-regulating effect could not be estimated with this methodology from the acquired data. It should be mentioned here that the afore-



(a) Decrease in active power at 20:45 Jul.07, 2019.



(b) Decrease in active power at 22:00 Jun.28, 2019.

Figure 4.3: Frequency and active power change during the time of $15\,\mathrm{min}$ products.

mentioned methodology, which successfully worked in the ’90s [65], cannot be fruitful nowadays, since the system is much larger and the effect of outages much smaller.

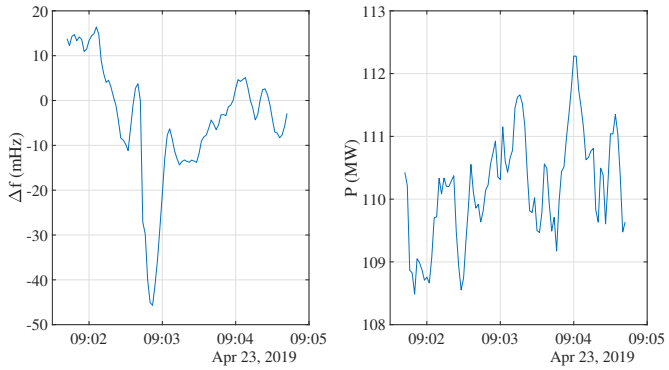


Figure 4.4: Frequency and active power change on Apr.23, 2019.

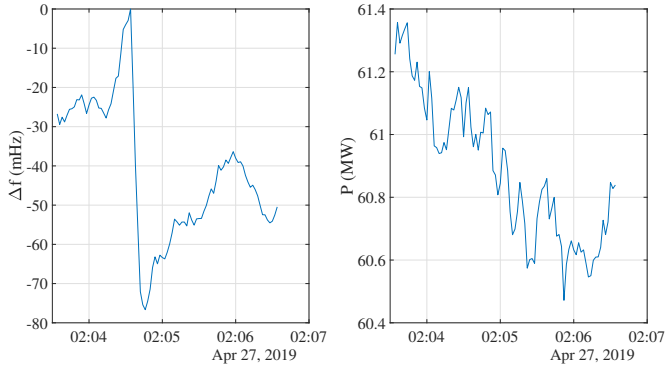


Figure 4.5: Frequency and active power change on Apr.27, 2019.

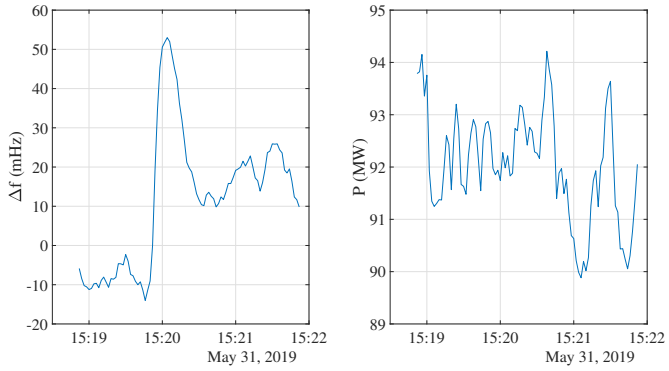


Figure 4.6: Frequency and active power change on May 31, 2019.

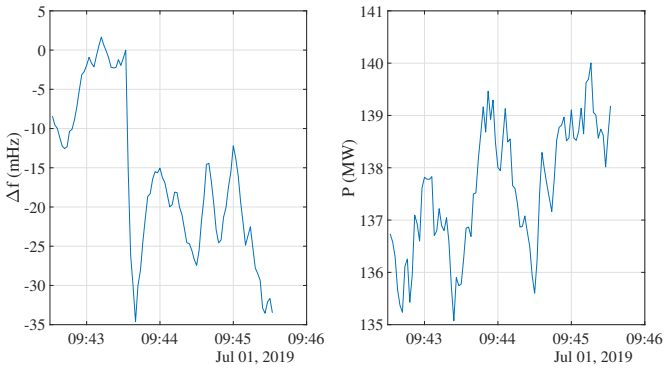


Figure 4.7: Frequency and active power change on Jul.01, 2019.

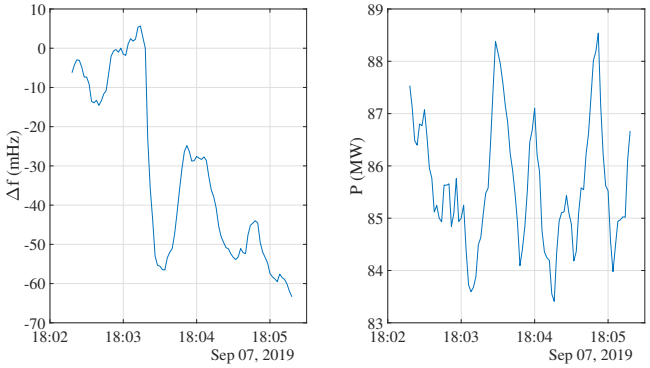


Figure 4.8: Frequency and active power change on Sep.07, 2019.

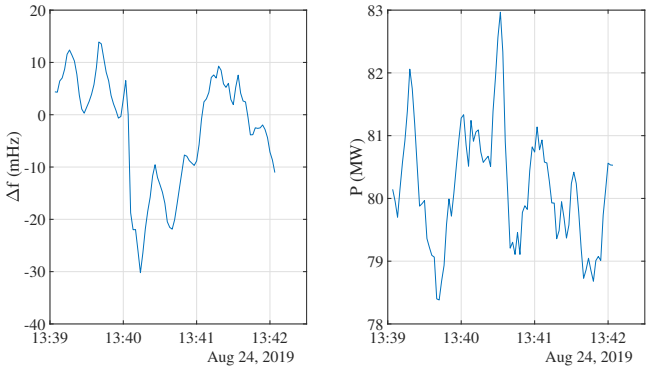


Figure 4.9: Frequency and active power change on Aug.24, 2019.

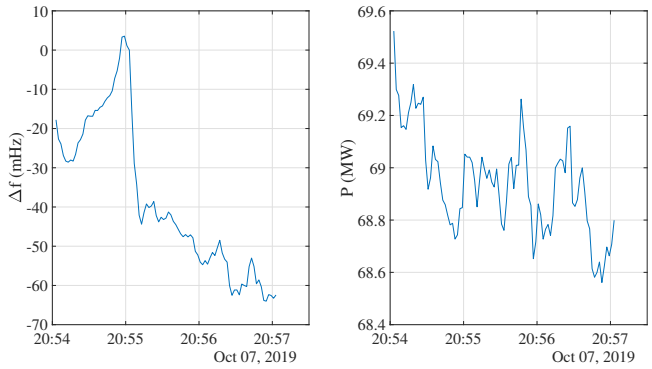


Figure 4.10: Frequency and active power change on Oct.07, 2019.

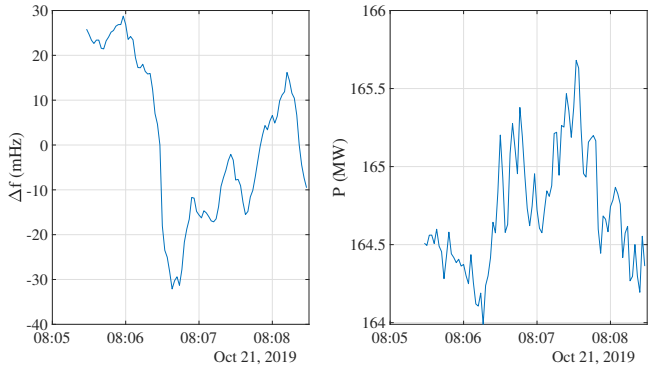


Figure 4.11: Frequency and active power change on Oct.21, 2019.

4.2 Deterministic frequency deviations

It was observed that deterministic frequency changes occur on a daily basis every 15 minutes, which basically correspond to the $15\,\mathrm{min}$ products of the electricity market. All the acquired measurements are shown in Fig. 4.14 including both weekdays and weekends. The colored curves denote the individual frequency measurements, while the black bold curve is the average product of all recordings. Importantly, the biggest frequency variations happen from 21:00 to 03:00, while remarkable frequency changes occur around 7:00 in the morning and 18:00 in the afternoon.

In order to obtain a better insight of the system dynamics, the frequency variations and the corresponding active power response are examined for every hour

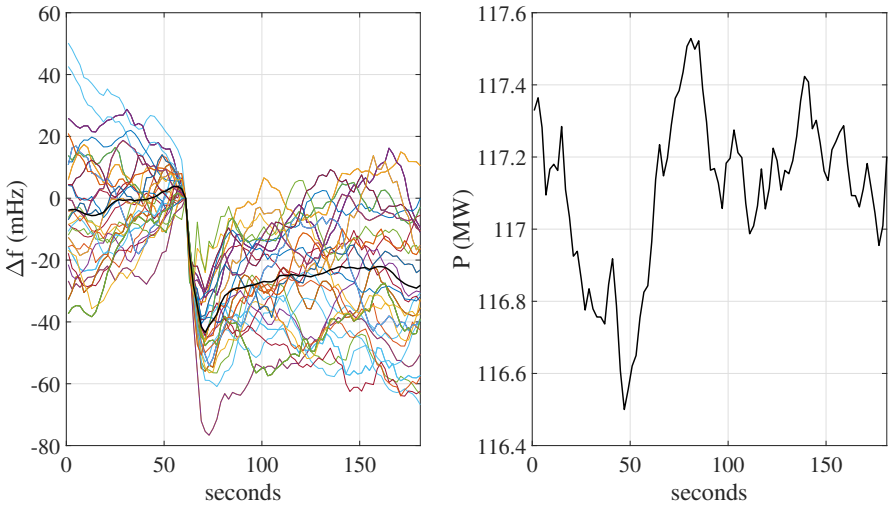


Figure 4.12: Average frequency and active power signals for frequency drops.

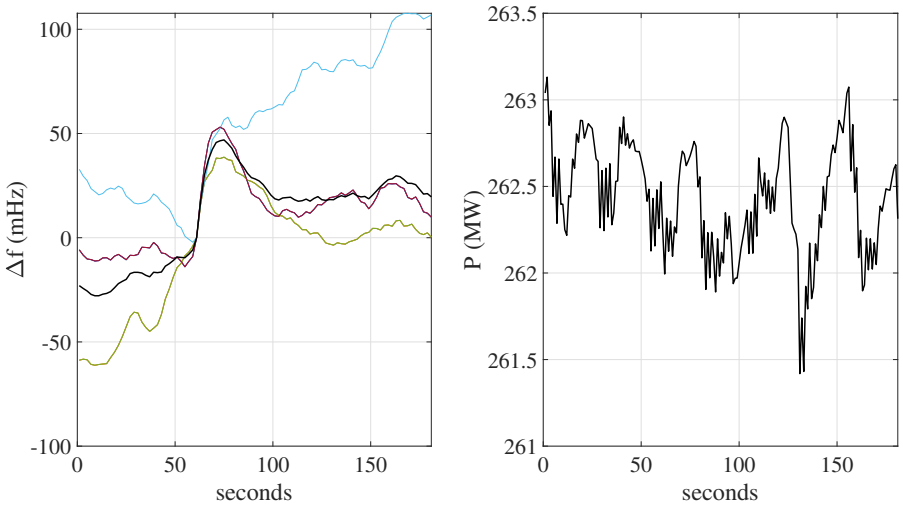


Figure 4.13: Average frequency and active power signals for frequency rises.

within a day. It should be also mentioned that, since the total power consumption significantly varies within a weekday and thus, the noise caused by the switching of the consumers is strong, only the measurements acquired in the weekends are employed. This phenomenon can be justified by the fact that apart from residen-

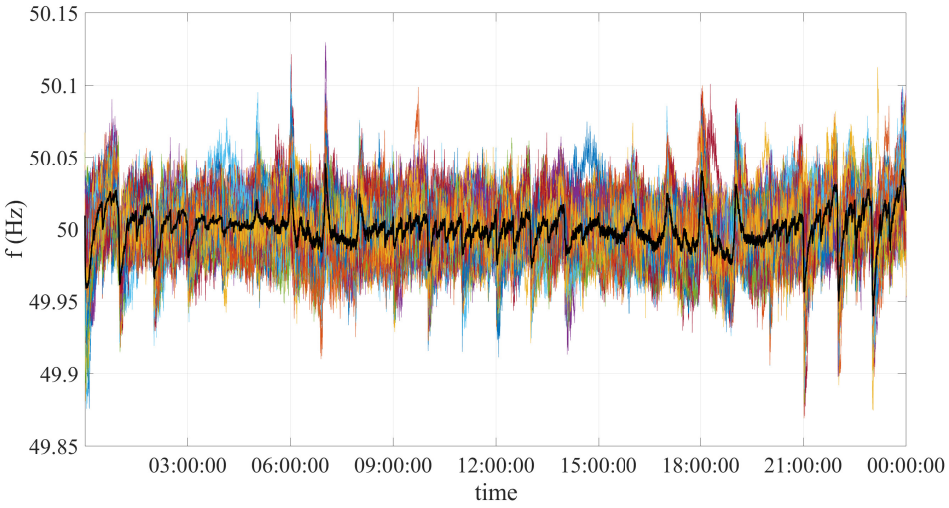


Figure 4.14: Daily frequency variations.

tial loads and commercial, industrial loads are also scheduled during the weekdays while they usually remain inactive during the weekends. Those extra loads lead to bigger active power changes during the time steps of the hourly market products. Those bigger active power changes overlap the self-regulating effect and therefore, its estimation is not feasible. On the contrary, weekends are characterized by smaller active power changes at the time steps of the hourly market products (due to the smaller load demand) and thus, the impact of overlapping is expected to be lower. Moreover, the hourly power consumption curves are characterized by consumption trends that should be removed in order to isolate the influence of the frequency upon the active power. An example of a power consumption trend is illustrated in Fig. 4.15, in which the active power linearly decreases throughout the examined time period. The yellow bold curve in the middle of the graph highlights the linear trend as calculated by averaging all the recordings of the active power.

To tackle this problem, a linear detrend was applied to the average power consumption of each given hour. Importantly, the average active power is deployed, since the individual measurements are noisy due to the continuous switching of the consumers. A time horizon of 10 minutes (5 minutes before and after the fre-

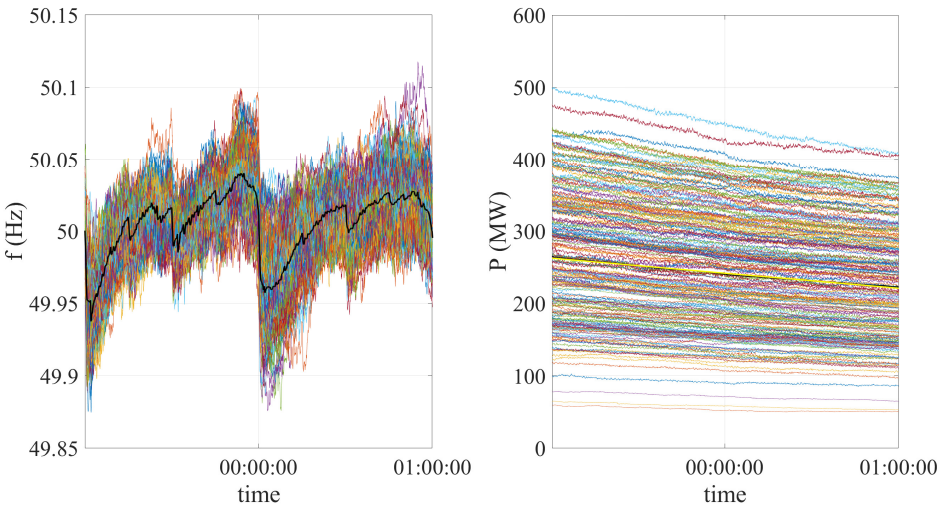


Figure 4.15: Linear trend in active power at midnight.

quency change) is used for presenting the plots in order to better observe the filtered active power response. The hours with significant frequency changes (deltas) are illustrated in Fig. 4.16-4.29. The left plots depict the individual frequency changes (colored curves) as well as their average (black bold curve). On the right side, the detrended average active power is illustrated, which describes the change in active power without the influence of consumption trends.

Throughout a day, common behavior was observed using the generated graphs. Importantly, during the early morning hours (00:00, 01:00, 02:00 and 04:00) the active power rises at the time that the frequency starts to decrease. Therefore, no self-regulating effect is visible during those hours. Similarly, at 03:00, 10:00-14:00, 18:00 and 21:00-23:00 the active power does not follow the frequency variations. In addition, the active power signal is very noisy and thus, no meaningful conclusion can be extracted. Furthermore, at 05:00-9:00, 15:00-17:00 and 19:00-20:00 the frequency does not change more than $20\,\mathrm{mHz}$. To conclude, no certain conclusion can be drawn with respect to the self-regulating effect, since the active power does not follow the corresponding deterministic frequency variations. Rather, the active power changes in this examined network group seem to be a (small) part of the

cause of the deterministic frequency deviations.

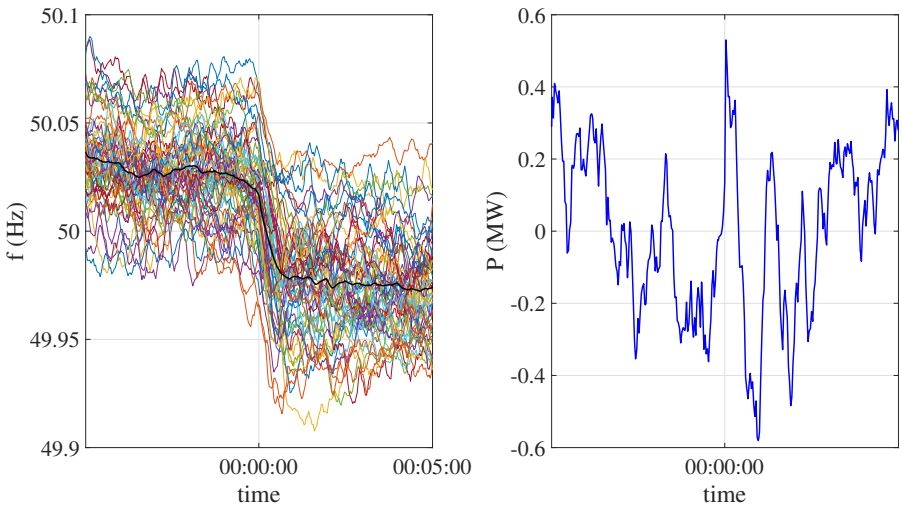


Figure 4.16: Weekend frequency variations at 00:00.

4.3 Summary

The aim of this chapter is to determine the steady-state relationship between frequency and active power, i.e., self-regulating effect of the grid. To do so, measurements from three EHV/HV substations powering a large grid were utilized. Two different approaches were followed; the first one is based on individual frequency changes that meet a specific criterion and the second one is based on the daily deterministic frequency deviations.

Regarding the first approach, several measurements were acquired following the presented frequency gradient criterion. However, the corresponding active power signals contain noise coming from the switching of the load and the changes in frequency are not followed by similar changes in active power. Therefore, individual measurements cannot be employed in practice, since the signal-to-noise ratio is rather low. In an attempt to increase the signal-to-noise ratio, an ensemble averaging of the active power signals have been performed. It was observed that there was

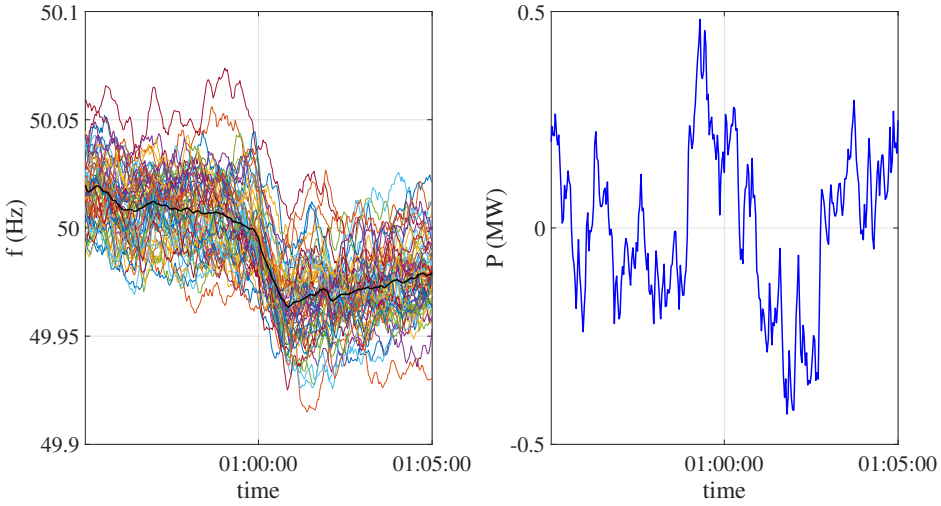


Figure 4.17: Weekend frequency variations at 01:00.

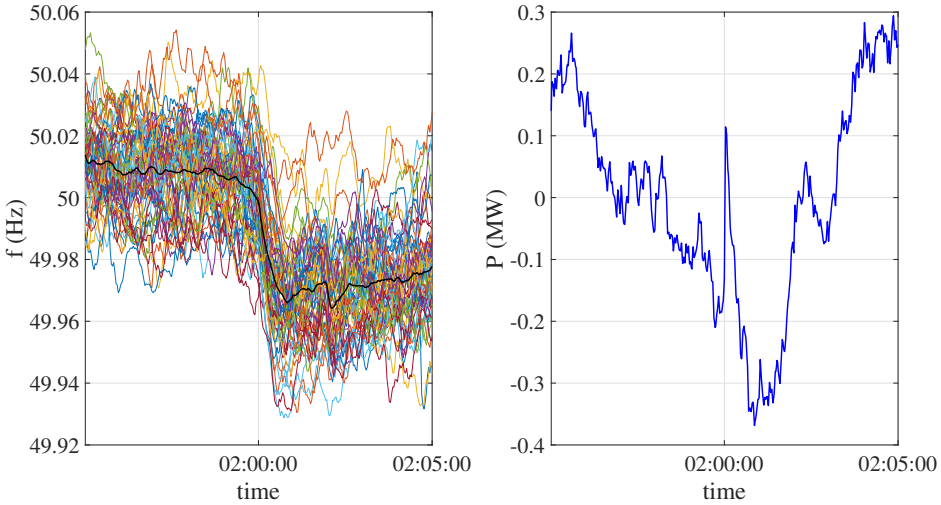


Figure 4.18: Weekend frequency variations at 02:00.

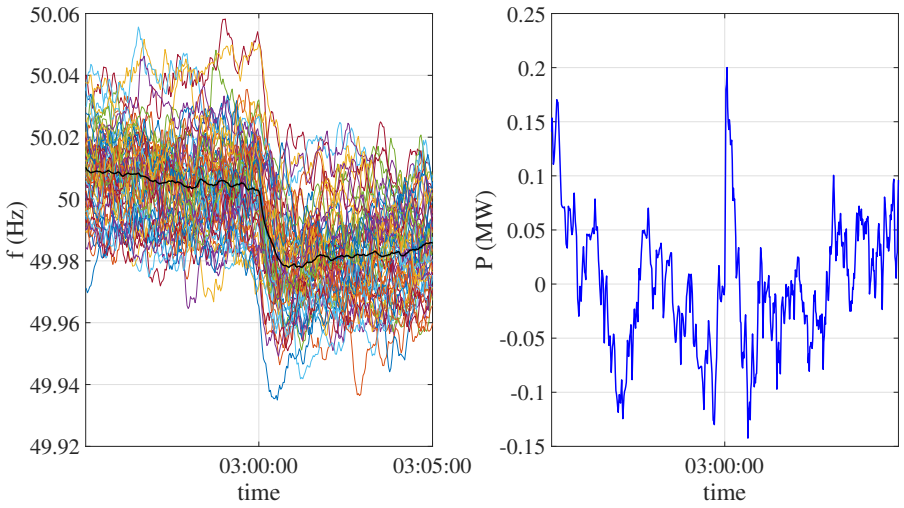


Figure 4.19: Weekend frequency variations at 03:00.

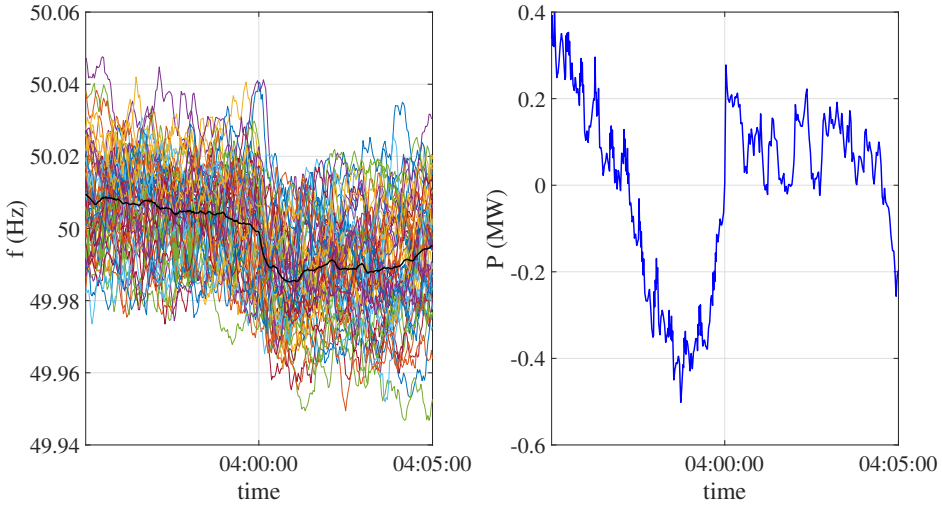


Figure 4.20: Weekend frequency variations at 04:00.

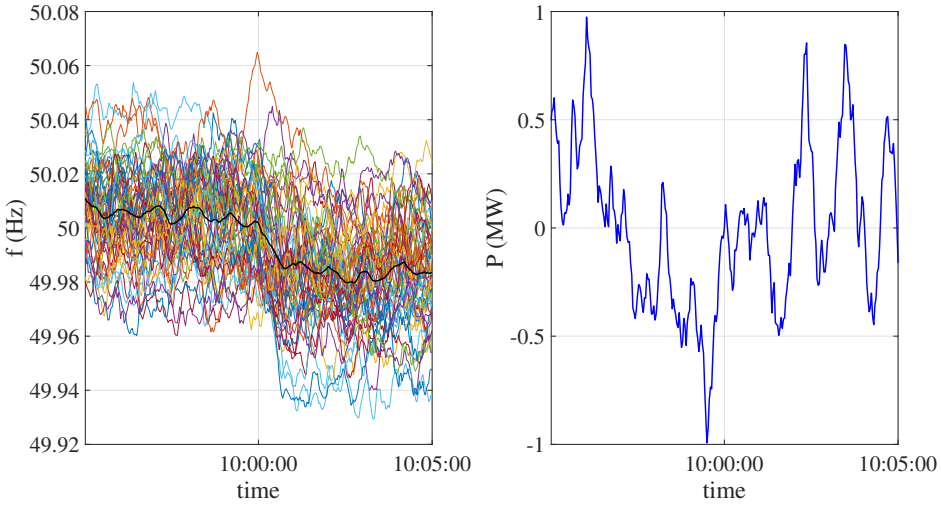


Figure 4.21: Weekend frequency variations at 10:00.

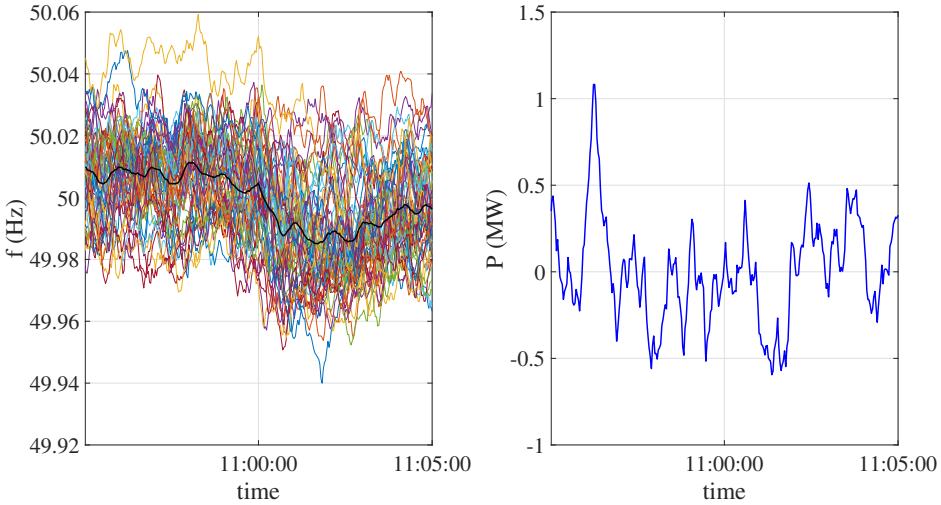


Figure 4.22: Weekend frequency variations at 11:00.

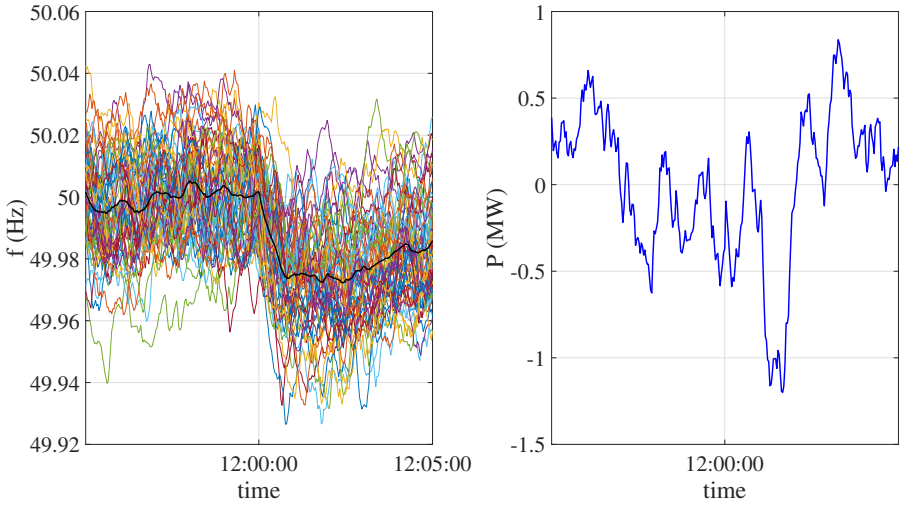


Figure 4.23: Weekend frequency variations at 12:00.

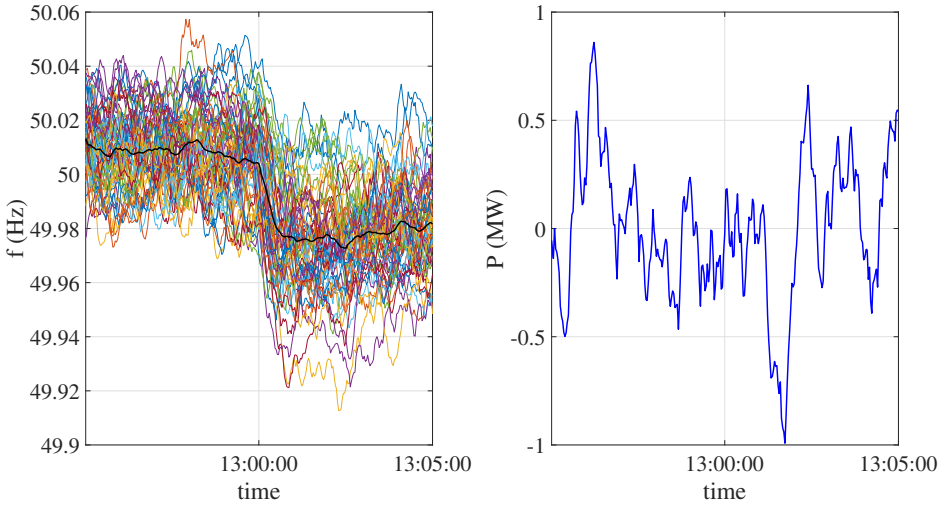


Figure 4.24: Weekend frequency variations at 13:00.

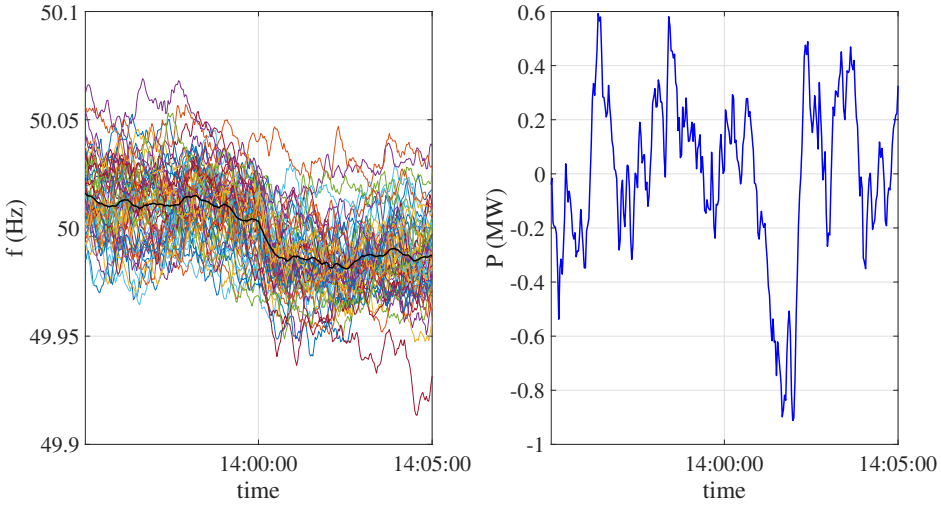


Figure 4.25: Weekend frequency variations at 14:00.

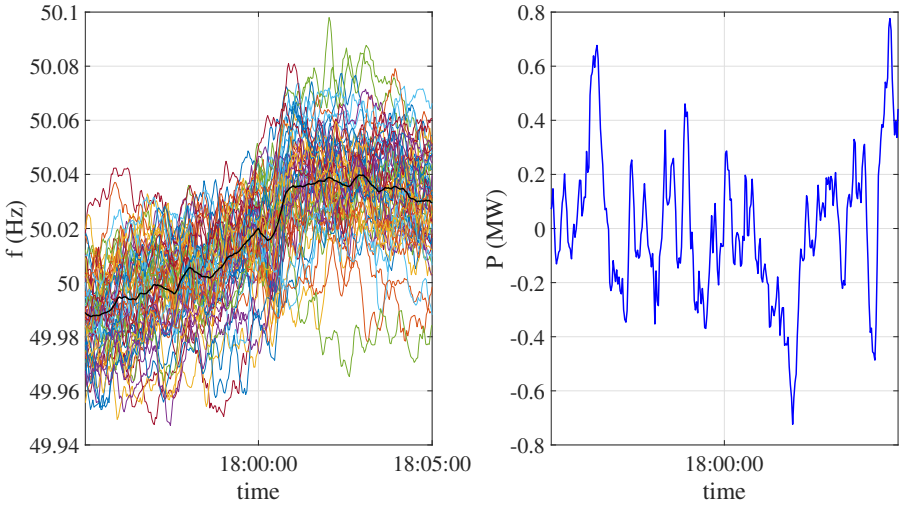


Figure 4.26: Weekend frequency variations at 18:00.

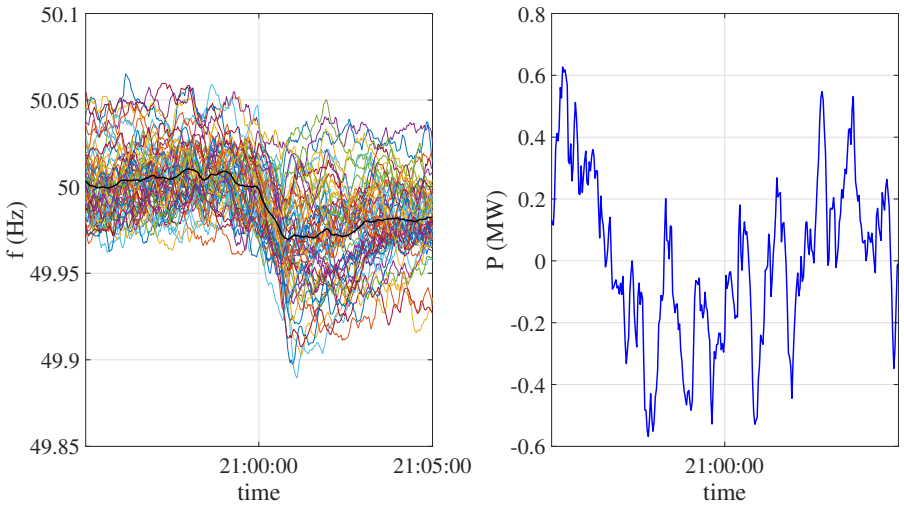


Figure 4.27: Weekend frequency variations at 21:00.

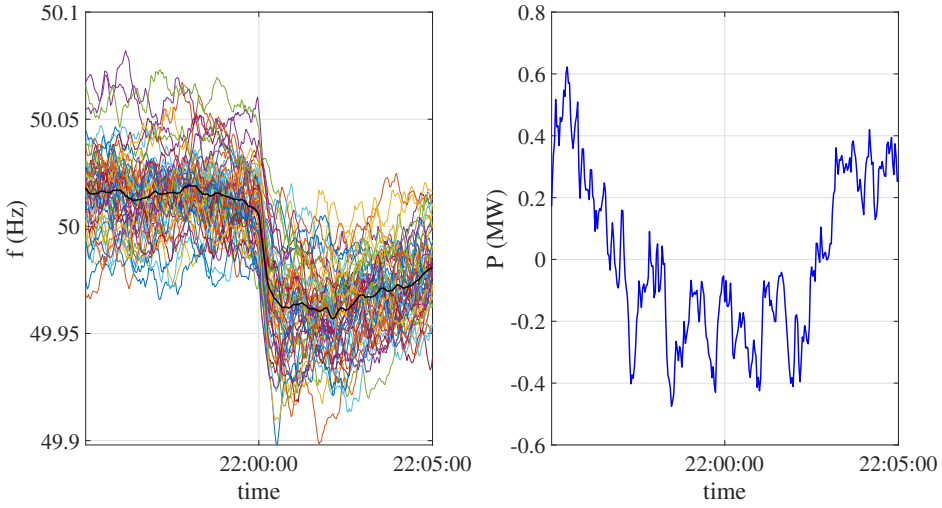


Figure 4.28: Weekend frequency variations at 22:00.

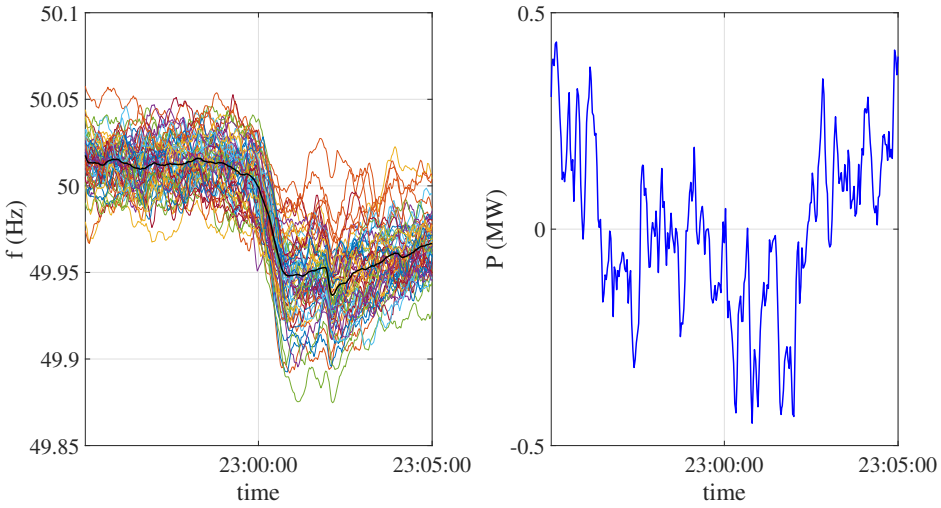


Figure 4.29: Weekend frequency variations at 23:00.

no remarkable change in the active power signal caused by the change in frequency and the stochastic nature of the load overlaps any self-regulating effect.

As for the second approach, deterministic frequency deviations occur on a daily basis every 15 minutes, corresponding to the 15 min. products of the electricity market. In order to remove the influence of electricity market products and the noise coming from the switching of the consumers, the weekend measurements were selected and then, the consumption trends were removed from the active power signals. Yet any self-regulating effect was overlapped mostly by noise.

As a suggestion for future work, measuring more ADNs simultaneously might lead to an increase in the signal-to-noise ratio. This would further lead to an increase in the probability of capturing the self-regulating effect at the cost of a lower spatial resolution of the results nonetheless. Another possible alternative would be measuring over longer periods so that large disturbances, e.g., system split, are captured.

Chapter 5

Identification of ADN Dynamics using Cluster Analysis

This chapter introduces many of the core contributions of this work. Those contributions can be also found published in [12, 115, 116]. In the beginning, it motivates the need for adequate data handling and preprocessing solutions in the measurement-based approach. Then, this chapter rigorously describes the proposed unsupervised learning method for clustering dynamic behavior that can be applied to any time-varying dynamical system characterized by distinct sets of dynamics, regardless of the science field. Next, the clustering results using the acquired field data are presented and discussed.

5.1 Motivation

References [25, 26, 39] have demonstrated that the load model parameters may strongly vary from hour to hour, day to day, and month to month. In this context, it is commonly recognized that models relying only on a small amount of measurements are unreliable and thus, measurements should be acquired over a reasonable amount of time so that seasonal grid variations are considered [24, 35, 39]. However, although measuring over, e.g., a year can capture a wide range of grid config-

urations, irrelevant measurements can be also recorded as the trigger criteria cannot distinguish the cause of the disturbance.

In this regard, since those dynamic models are expected to be deployed in power system dynamic simulations and stability studies, data that do not reflect the response of the ADN to an external disturbance should be discarded from the system identification procedure [2]. For example, faults within the examined grid, wrongly triggered measurements, connection/disconnection of big consumers/DG within the examined grid, highly noisy data and outliers may be additionally recorded. Hence, long measuring periods usually result in big datasets, in which manual data processing would be an extremely time-demanding task. Besides that, identifying a different model for each measurement will simply yield overfitted models, valid only for a specific grid configuration. To this end, the optimal number of dynamic models, which adequately capture the general ADN characteristics while maintaining reliable levels of accuracy is an open question and strongly depends on the recorded dataset.

Based on the aforementioned ADN modeling issues, we propose a three-stage modeling approach which develops the concept of generating a few characteristic models for the same ADN in order to accurately describe the system under a seemingly infinite number of states, i.e., load/DG amount and composition. At the first two stages, we introduce a novel clustering methodology in order to firstly, detect irrelevant measurements and discard them from the modeling procedure and secondly, group the remaining data into clusters with similar dynamic behavior. At the third stage, we build a generic dynamic equivalent for each one of the derived clusters using the respective grouped entries. To do so, three different dynamic models are proposed and are thoroughly described in Chapter 6, 7, and 8. Finally, it should be pointed out that, although the proposed method was principally designed to work with field measurements and their peculiarities, it can be easily run with synthetic data generated by simulations of detailed ADN models.

5.2 Proposed methodology

The three-stage proposed method can be described by the flowchart in Fig. 5.1. As already mentioned before in the thesis, a dataset $\mathcal{D}$ of M measurements is recorded at the point of common coupling (PCC). Each measurement $i \in \{1, ..., M\}$ comprises three equally sized vectors $\boldsymbol{V}_i$, $\boldsymbol{P}_i$, and $\boldsymbol{Q}_i$ corresponding to the RMS values of the time domain (TD) signals of voltage, active, and reactive power, respectively. Note that this work examines the dynamic behavior of an ADN under electromechanical transients. During those transients, the bus voltage may vary by $\pm 10\%$ from the nominal value [43], and thus the acquired measurements should contain voltage disturbances within those limits.

As a first step (stage 0), the recorded data are preprocessed in order to create a common reference for all the measurements. Based on the resulting signals, the clustering features are extracted and normalized. Then (stage 1), all the irrelevant data are identified and discarded using one of the proposed clustering algorithms, resulting in a new dataset $\mathcal{D}_r \subseteq \mathcal{D}$. In the second stage, similar dynamics are automatically grouped together using one of the proposed clustering algorithms over $\mathcal{D}_r$. In the final stage, the cluster entries are randomly split into training and validation set. The training set is used to derive one generic model for each cluster using system identification techniques while the accuracy and the generalization capability of the yielded models are assessed using the validation set.

5.2.1 Data preprocessing

A major characteristic differentiating the traditional distribution networks from ADNs is the installed DG. Therefore, the measured power values at the PCC do not reflect the real load demand. Instead, they express the difference between the real power demand and the power generated by DG. However, accessing detailed DG data in order to estimate the real load demand may be difficult, since there are countries where DSOs do not share those data with TSOs due to confidentiality rules [11]. To this end, a preprocessing phase is performed, as described in Chapter 3. Neverthe-

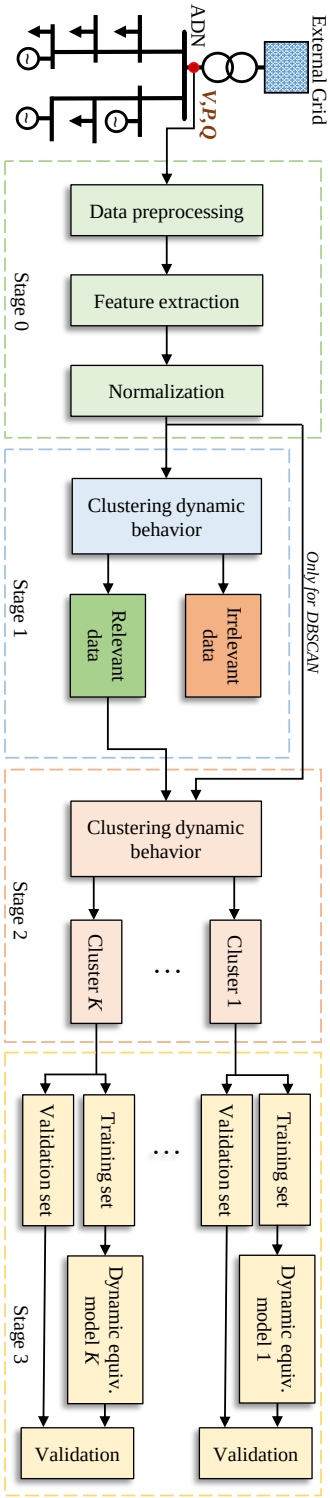


Figure 5.1: Flowchart of the proposed methodology.

less, for the sake of completeness, it is repeated in this section. During this step, the initial pre-disturbance conditions $V_{0,i}$, $Y_{0,i}$ are subtracted from the TD values $V_i(t)$, $Y_i(t)$ for voltage and power, respectively, using:

$$\Delta V_i(t) = V_i(t) - V_{0,i} \quad \text{and} \quad \Delta Y_i(t) = Y_i(t) - Y_{0,i}. \tag{5.1}$$

The variable Y can describe either the active or reactive power while $(V_{0,i}, Y_{0,i})$ can be considered any operating point within a few time steps before the disturbance. This step is fundamental for the generalization capability of the proposed method, since it virtually decouples the dynamic response of the system from its initial unknown power conditions and additionally yields a common reference for the cluster analysis.

5.2.2 Feature extraction and normalization

Clustering dynamic behavior is a rather unexplored topic in control theory. Consequently, the selection of the adequate clustering features for clustering dynamic behavior is not straightforward. A simple use of the recorded dynamic responses, as in clustering load profiles, would not work, since the model output depends also on the applied input. To this end, a novel methodology is proposed that can be readily generalized to any dynamical system characterized by distinct sets of time-varying dynamics in order to identify groups of similar dynamic behavior based only on observed input-output sequences.

The proposed clustering method relies on the principle that, in order to correctly compare multiple dynamical systems, it is essential to examine their responses under the same input. In this context, a unit step function u is one of the most commonly used signals for examining the response of a system in control theory due to its practicality and its rich frequency spectrum. However, in order to obtain the unit step response, the system should be first identified. To do so, a discrete transfer

function $H_i(z)$ is estimated for each measurement i as

$$\begin{aligned} H_i^n(z) = \frac{\mathcal{Z}\{\Delta Y_i(t)\}}{\mathcal{Z}\{\Delta V_i(t)\}} = \frac{\Delta Y_i(z)}{\Delta V_i(z)} = \\ = \frac{\beta_{0,i} + \beta_{1,i} z^{-1} + ... + \beta_{n,i} z^{-n}}{1 + \alpha_{1,i} z^{-1} + ... + \alpha_{n,i} z^{-n}}, \end{aligned} \tag{5.2}$$

where $\mathcal{Z}$ denotes the z-transform, n indicates the model order and

$$\boldsymbol{\theta}_i = [\alpha_{1,i}, ..., \alpha_{n,i}, \beta_{0,i}, ..., \beta_{n,i}]^\top \tag{5.3}$$

is a vector containing the transfer function parameters. Eq. (5.2) can be easily re-formulated as

$$\begin{aligned} \Delta Y_i(z) \left(1 + \alpha_{1,i} z^{-1} + ... + \alpha_{n,i} z^{-n}\right) = \\ = \Delta V_i(z) \left(\beta_{0,i} + \beta_{1,i} z^{-1} + ... + \beta_{n,i} z^{-n}\right). \end{aligned} \tag{5.4}$$

Subsequently, the z-domain (5.4) can be directly converted to a difference equation as

$$\begin{aligned} \Delta Y_i[t] = &- \alpha_{1,i} \Delta Y_i[t-1] - ... - \alpha_{n,i} \Delta Y_i[t-n] + \\ &+ \beta_{0,i} \Delta V_i[t] + ... + \beta_{n,i} \Delta V_i[t-n]. \end{aligned} \tag{5.5}$$

Taking into consideration all the T recorded samples, we can expand (5.5) to a general matrix form as

$$\overbrace{\begin{bmatrix} -\Delta Y_i[t-1] & \dots & \Delta V_i[t-n] \\ \vdots & \ddots & \\ -\Delta Y_i[t-1-T] & \dots & \Delta V_i[t-n-T] \end{bmatrix}}^{\boldsymbol{\Psi}_i} \overbrace{\begin{bmatrix} \alpha_{1,i} \\ \vdots \\ \beta_{n,i} \end{bmatrix}}^{\boldsymbol{\theta}_i} = \overbrace{\begin{bmatrix} \Delta Y_i[t] \\ \vdots \\ \Delta Y_i[T] \end{bmatrix}}^{\boldsymbol{Y}_i} \tag{5.6}$$

or in a more compact form:

$$\boldsymbol{\Psi}_i \cdot \boldsymbol{\theta}_i = \boldsymbol{Y}_i. \tag{5.7}$$

Solving for $\boldsymbol{\theta_i}$ using the linear least squares, we obtain

$$\boldsymbol{\theta}_i^* = \left(\boldsymbol{\Psi}_i^\top \boldsymbol{\Psi}_i\right)^{-1} \boldsymbol{\Psi}_i^\top \boldsymbol{Y}_i. \tag{5.8}$$

To identify the model order, the root mean square error

$$RMSE_i = \sqrt{\frac{1}{T}\sum_{t=t_0}^{T}\left(\Delta\hat{Y}_i(t) - \Delta Y_i(t)\right)^2} \tag{5.9}$$

between the model output $\Delta\hat{Y}_i$ and recorded values ΔY_i is employed. Starting from first order, the order of each transfer function keeps increasing until the change of RMSE between two consecutive model orders is not greater than a user-defined threshold ϵ_y. This iterative procedure results in two transfer functions $H^*_{P,i}(z|\boldsymbol{\theta}^*_i)$ and $H^*_{Q,i}(z|\boldsymbol{\theta}^*_i)$ that model the active and reactive power response to the i-th disturbance, respectively. Then, a unit step is applied to each of the derived models generating M step responses for active and reactive power, $\hat{y}_{P,i}$ and $\hat{y}_{Q,i}$, respectively. Since the signals $\hat{y}_{P,\{1,...,M\}}$ and $\hat{y}_{Q,\{1,...,M\}}$ were generated under the same input, we can leverage them to cluster dynamic behavior using their TD and frequency domain (FD) values as features. To do so, we calculate the frequencies at which the Fast Fourier Transform (FFT) of the TD signals $\hat{y}_{P,i}(t)$ and $\hat{y}_{Q,i}(t)$ reach their maximum amplitude, as:

$$f_{P,i} = \arg\max_f \text{FFT}\left(\hat{y}_{P,i}(t)\right), \tag{5.10}$$

$$f_{Q,i} = \arg\max_f \text{FFT}\left(\hat{y}_{Q,i}(t)\right). \tag{5.11}$$

Then, for every measurement i, we merge the generated TD signals $\hat{y}_{P,i}(t)$ and $\hat{y}_{Q,i}(t)$ with the frequencies $f_{P,i}$ and $f_{Q,i}$ into a single feature vector $\boldsymbol{F}_i$ as:

$$\begin{aligned}\boldsymbol{F}_i = [&\hat{y}_{P,i}(t_0), ..., \hat{y}_{P,i}(T), \hat{y}_{Q,i}(t_0), ..., \hat{y}_{Q,i}(T),\\ &f_{P,i}, ..., f_{P,i}, f_{Q,i}, ..., f_{Q,i}]^\top.\end{aligned} \tag{5.12}$$

In order to balance the influence of TD and FD features, the frequencies $f_{P,i}$ and $f_{Q,i}$ are expanded within the feature vector so that they occupy the same number of vector positions as their TD signals.

Once all individual feature vectors have been formed, they are put into a matrix $\boldsymbol{F} = [\boldsymbol{F}_1, \boldsymbol{F}_2, ..., \boldsymbol{F}_M]$. Then, $\boldsymbol{F}$ is scaled using the Min-Max normalization technique, so that all matrix elements are within $[0, 1]$ [117]. Finally, the scaled $\boldsymbol{F}$ is

used as an input to the clustering algorithm. A high-level description of the steps followed for clustering dynamic behavior are shown in the pseudocode of Algorithm 1 below.

Algorithm 1 Pseudocode for clustering dynamic behavior.

for $i = 1 : M$ and for both active and reactive power **do**
  $n = 1$
  Calculate $H_i^1(z)$
  Calculate RMSE_i^1
  repeat
    $n = n + 1$
    Calculate $H_i^n(z)$
    Calculate RMSE_i^n
  until $\left(\text{RMSE}_i^n - \text{RMSE}_i^{n-1}\right) < \epsilon_y$
  Calculate unit step response
  Calculate max. frequency of the unit step response FFT
  Form $\boldsymbol{F}_i$
end for
Merge individual $\boldsymbol{F}_i$ to a feature matrix $\boldsymbol{F}$
Normalize $\boldsymbol{F}$
Run the clustering algorithm with $\boldsymbol{F}$ as input

It is worth pointing out that we avoid performing the clustering based on the parameters of existing load/ADN models, since the clustering procedure would be highly dependent on the deployed model and the accuracy of the parameter estimation. Furthermore, complex models with a lot of parameters require nonlinear optimization which may result in local optimum solutions [24]. Hence, multiple combinations of parameters might describe the same dynamics, and thus using them as features would result in redundant clusters. On the contrary, this clustering approach relies exclusively on the recorded dynamic responses and thus, it can be effortlessly integrated into most of the existing load modeling approaches in the literature.

5.2.3 Clustering dynamic behavior

In this phase, measurements that correspond to similar dynamic behavior are automatically grouped together. One of the proposed clustering algorithms is initially applied to $\mathcal{D}$ (stage 1) and then to $\mathcal{D}_r$ (stage 2).

Since clustering dynamic behavior has, to the author's best knowledge, not been presented in the literature before, the choice of a suitable clustering method was not straightforward. Hence, four algorithms of different notion and complexity, which are based on four different types of clustering methods (partitioning, hierarchical, fuzzy, density-based), were implemented and validated; namely k-means++ [118], agglomerative hierarchical clustering [119], fuzzy c-means [120] and DBSCAN [121]. It should be noted that DBSCAN does not require a two-stage run since it discards outliers while clustering the data.

Here, the theoretical foundations of the deployed clustering algorithms are briefly discussed.

1. ***k*-means++**: This algorithm splits the data into K clusters $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_K\}$ by repeatedly minimizing the "Within Cluster Sum of Squares" (WCSS) [118]. This minimization is described by:

$$\arg\min_{\mathcal{C}} \sum_{k=1}^{K} \sum_{\boldsymbol{F}_i \in \mathcal{C}_k} \|\boldsymbol{F}_i - \boldsymbol{c}_k\|^2, \tag{5.13}$$

   where

$$\boldsymbol{c}_k = \frac{1}{N_k} \sum_{\boldsymbol{F}_i \in \mathcal{C}_k} \boldsymbol{F}_i \tag{5.14}$$

   denotes the centroid of $\mathcal{C}_k$, as calculated by its N_k entries.

2. **Agglomerative hierarchical clustering**: This algorithm starts with the hypothesis that every entry forms its own cluster. Then, a similarity metric, e.g., euclidean distance, is calculated between the clusters and is stored in an $M \times M$ symmetrical matrix. As a next step, the two most "similar" clusters are merged together and a new $(M-1) \times (M-1)$ similarity matrix is estimated. The same procedure is repeated until the desired number of clusters is

reached [119]. In our implementation, we deploy the euclidean distance as a similarity metric.

3. **Fuzzy c-means**: This method is based on the simple principle that each data entry can partially belong to more than one cluster. In a iterative manner, fuzzy c-means tries to split the data into clusters by assigning the adequate degrees of membership and minimizing the sum of squares [120]:

$$J_m = \sum_{i=1}^{M} \sum_{k=1}^{K} \mu_{ik}^m \|\boldsymbol{F}_i - \boldsymbol{c}_k\|^2, \tag{5.15}$$

where μ_{ik} denotes the degree of membership of $\boldsymbol{F}_i$ in the k-th cluster, while m expresses the fuzzy partition matrix exponent which regulates the degree of fuzzy overlap. In this implementation, m is set to 2 and since the goal is to build a different dynamic model for each cluster, each measurement is assigned to the cluster with the maximum membership after the algorithm has converged to a solution.

4. **DBSCAN**: The Density-Based Spatial Clustering of Applications with Noise is an algorithm suitable to discover clusters with arbitrary shape without requiring *a priori* the total number of clusters as input [121]. The algorithm classifies each point as core, border or noise point based on how many neighbor points are found within a distance ϵ. The minimum number of neighbor points minPts and the distance ϵ are the two user-defined parameters of the algorithm. A core point has at least minPts points within ϵ, a border point has less than minPts as neighbors but at least one core point within ϵ and the noise points have no neighbor points. Finally, the neighbor core and border points are grouped to the same cluster. In this application, the minPts is set to 10 and the ϵ is set to one of $\{0.2, 0.3, 0.4\}$ based on the examined dataset.

To determine the optimal number of clusters in stage 1, the gap statistic is deployed since it usually outperforms other popular methods, e.g., the "elbow" [122].

The gap statistic essentially describes the difference of the intra-cluster variation

$$W_k = \sum_{k=1}^{K} \frac{1}{2\|\boldsymbol{c}_k\|} \sum_{i,j \in \mathcal{C}_k} \|\boldsymbol{F}_i - \boldsymbol{F}_j\|^2 \tag{5.16}$$

with its expectation under a uniform distribution of the data E^* for a varying number of clusters k. Its analytic expression can be written as

$$\text{Gap}(k) = E^*[\log(W_k)] - \log(W_k), \tag{5.17}$$

while the optimal k can be defined by

$$k^* = \arg\max_k \text{Gap}(k). \tag{5.18}$$

In stage 2, once the irrelevant clusters have been removed, a visual inspection of the remaining clusters is performed in order to identify the dominant ADN dynamics. The user can exploit this information to set the number of clusters of stage 2. Theoretically, the gap statistic could be used in stage 2 as well. However, since the optimal number of clusters is highly dependent on the application requirements and commonly used metrics provide only an estimate of k^*, the deployment of any of those metrics is avoided. Instead, the user is free to define the number of clusters based on his/her modeling requirements.

5.3 Clustering results

The aim of this section is to highlight the existence of different dynamics within an ADN, which cannot be neglected during the modeling procedure. The section is split into two parts; the clustering results of stage 1 and stage 2, respectively.

5.3.1 Stage 1

Although clustering algorithms have been widely used in the context of power systems [123], their application in clustering dynamic behavior has not been documented. Therefore, we evaluate the four selected algorithms using the metrics of

"within cluster sum of squares to between cluster variation" (WCBCR), the Davies-Bouldin index (DBI) and the average Silhouette index [124, 125]. Those metrics are mathematically described by the following equations:

$$\text{WCBCR} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{M_k} d^2(\boldsymbol{c}_k, \boldsymbol{F}_i)}{\sum_{1\leq k_1\leq k_2}^{K} d^2(\boldsymbol{c}_{k_1}, \boldsymbol{c}_{k_2})}, \tag{5.19}$$

where M_k denotes the total number of entries assigned to cluster $\mathcal{C}_k$ and $d(\bullet, \bullet)$ denotes the euclidean distance. In the numerator, WCBCR expresses the sum of the distances of all data entries to their center while the sum of distances between cluster centers forms the denominator.

DBI expresses an average similarity of each cluster with each most similar one and can be written as:

$$\text{DBI} = \frac{1}{K}\sum_{k=1}^{K}\max_{k_1\neq k_2}\left[\frac{\hat{d}(\mathcal{C}_{k_1}) + \hat{d}(\mathcal{C}_{k_2})}{d(\boldsymbol{c}_{k_1}, \boldsymbol{c}_{k_2})}\right], \tag{5.20}$$

where

$$\hat{d}(\mathcal{C}_k) = \sqrt{\frac{1}{2M_k}\sum_{i=1}^{M_k} d^2(\boldsymbol{c}_k, \boldsymbol{F}_i)}. \tag{5.21}$$

The Silhouette index is a metric for assessing how well an entry is fitted to its assigned cluster. To do so, this metric uses the following two functions:

$$a(\boldsymbol{F_i}) = \frac{1}{|\mathcal{C}_{k_1}| - 1}\sum_{j\in\mathcal{C}_{k_1}, i\neq j} d(\boldsymbol{F}_i, \boldsymbol{F}_j) \tag{5.22}$$

and

$$b(\boldsymbol{F_i}) = \min_{\mathcal{C}_{k_1}\neq\mathcal{C}_{k_2}}\frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_{k_2}} d(\boldsymbol{F}_i, \boldsymbol{F}_j). \tag{5.23}$$

Function $a(\boldsymbol{F_i})$ describes the average dissimilarity of the entry i to all its cluster members whereas function $b(\boldsymbol{F_i})$ expresses the minimum average dissimilarity of the entry i over the clusters that it is not assigned to. By combining these two function, the Silhouette index of the entry i is defined as:

$$s(\boldsymbol{F_i}) = \frac{b(\boldsymbol{F_i}) - a(\boldsymbol{F_i})}{\max(a(\boldsymbol{F_i}), b(\boldsymbol{F_i}))}. \tag{5.24}$$

In our application, we deploy the average Silhouette index over all entries and clusters. As expected, WCBCR and DBI need to be as low as possible, whereas the Silhouette index should be 1 in the best-case scenario.

The performance of the algorithms in stage 1 of the proposed methodology is evaluated by estimating the aforementioned metrics for different numbers of clusters ranging from two to ten. Since DBSCAN automatically discards the irrelevant measurements and does not partition the data based on a predefined number of clusters, its performance is assessed in stage 2. Importantly, the deployed metrics yielded comparable results for all the substations. Indicative results of six ADNs are

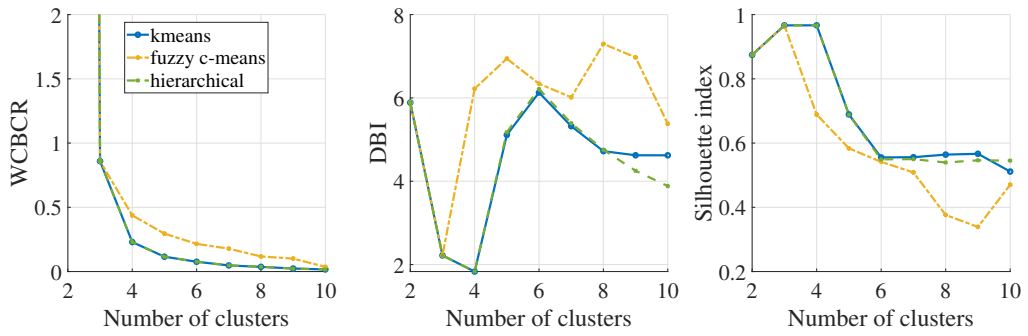


Figure 5.2: Sub.A.1: WCBCR, DBI, and Sihlouette index for different clustering algorithms.

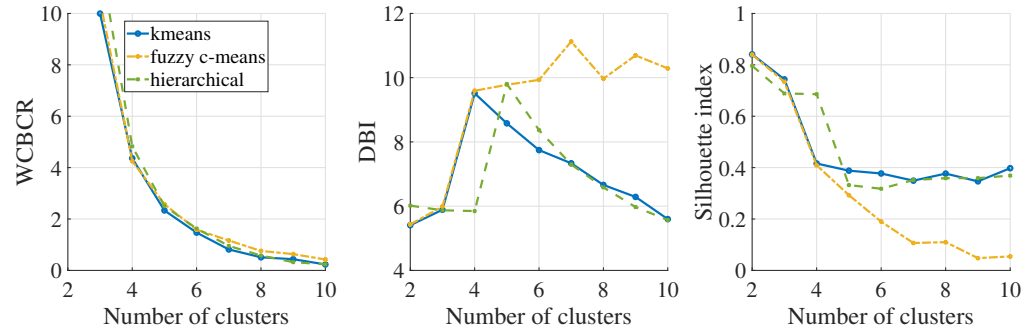


Figure 5.3: Sub.B.1: WCBCR, DBI, and Sihlouette index for different clustering algorithms.

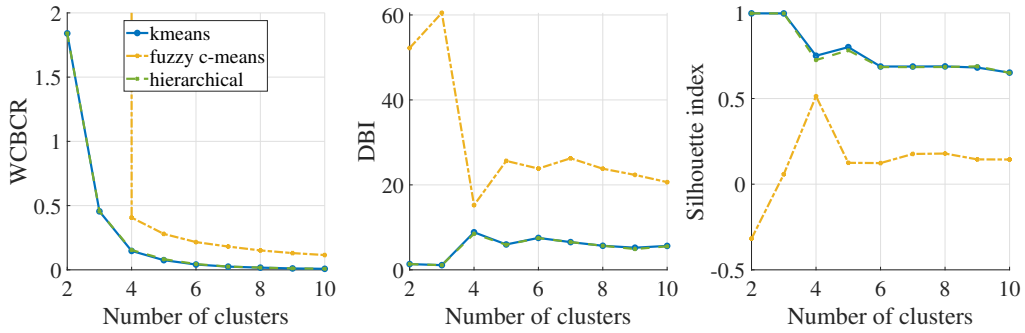


Figure 5.4: Sub.C.1: WCBCR, DBI, and Sihlouette index for different clustering algorithms.

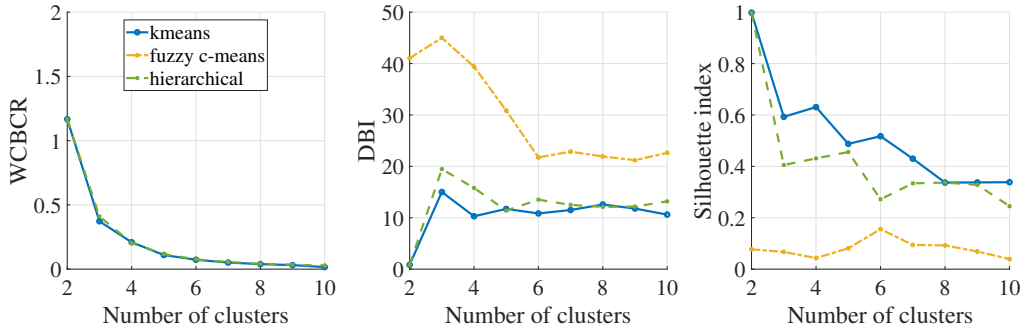


Figure 5.5: Sub.C.2: WCBCR, DBI, and Sihlouette index for different clustering algorithms.

illustrated in Fig. 5.2-5.7. It is clear that k-means++ and agglomerative hierarchical show similar results in every metric while both of them outperform fuzzy c-means. This result may be partially justified by the fact that each entry was assigned to the cluster with the maximum degree of membership even though fuzzy c-means partitions the dataset based on the notion of multiple memberships [120].

In general, the ability of a load/ADN model to effectively handle bad or irrelevant data for the parameter identification is particularly important [2, 18, 126, 127]. Since a major scope of those dynamic equivalent models is their integration to power system stability studies [2], the measurement data deployed for the parameter

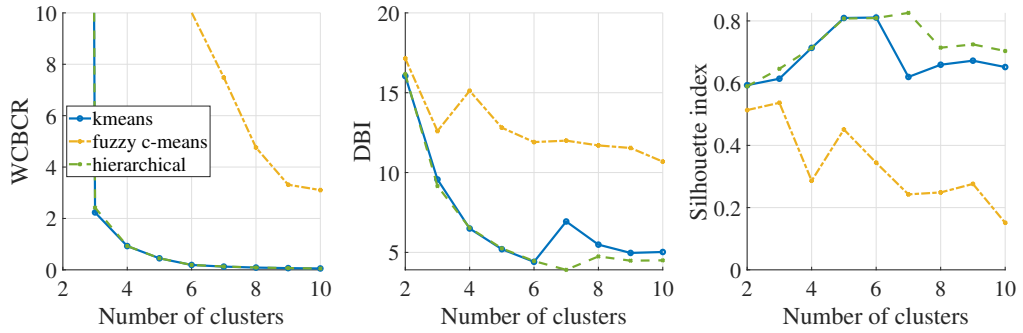


Figure 5.6: Sub.D.1: WCBCR, DBI, and Sihlouette index for different clustering algorithms.

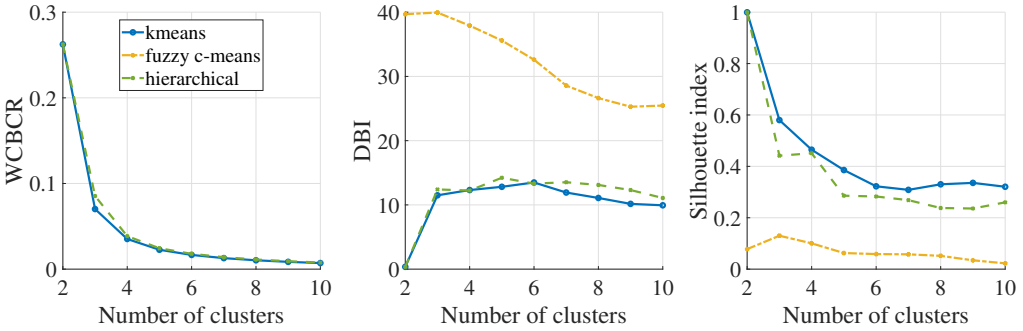


Figure 5.7: Sub.D.2: WCBCR, DBI, and Sihlouette index for different clustering algorithms.

estimation should reflect a clear response of the ADN to variations of the transmission grid voltage. Therefore, measurement data representing the response of the ADN to internal events are irrelevant for the purpose of the model. Those data are usually identified by the opposite direction between the change in voltage and power at the moment of the disturbance. Moreover, signals with noise levels comparable to the dynamic response of the system have a negative impact on the parameter estimation and should be discarded from the identification procedure [43]. Finally, clusters with a total number of entries below a certain threshold (1% of the total number of measurements acquired) can be considered as outliers and are omitted

from the clustering of stage 2, as the aim of the proposed methodology is to build generic models that are valid for a wide range of operating conditions.

Here, it is meaningful to illustrate a representative set of discarded clusters generated in this stage using the actual differences of voltage, active, and reactive power in pu. In Fig. 5.8-5.14, the grey curves depict the individual measurements while the black bold curves represent the average of the respective measurements. For example, the cluster in Fig. 5.13 contains measurements with some high frequency oscillations in active and reactive power which were present before the disturbance and do not correspond to the response of an ADN after a voltage change. Similar results but with a lower frequency are also depicted in Fig. 5.12, justifying the

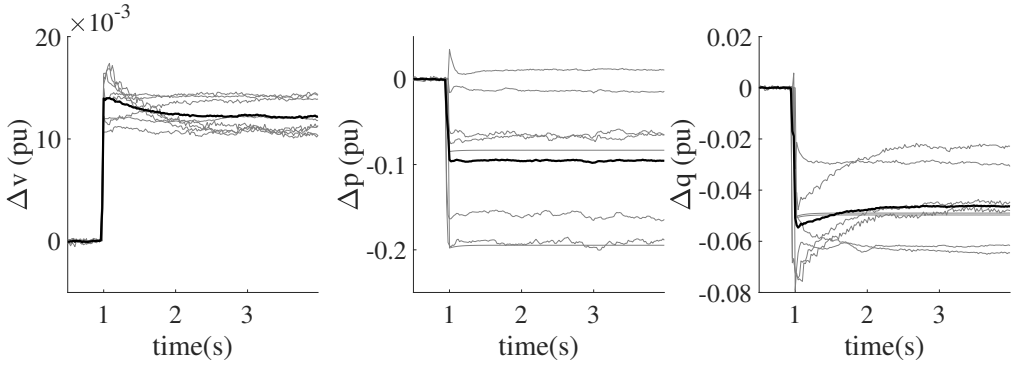


Figure 5.8: Sub.A.1: Indicative discarded cluster.

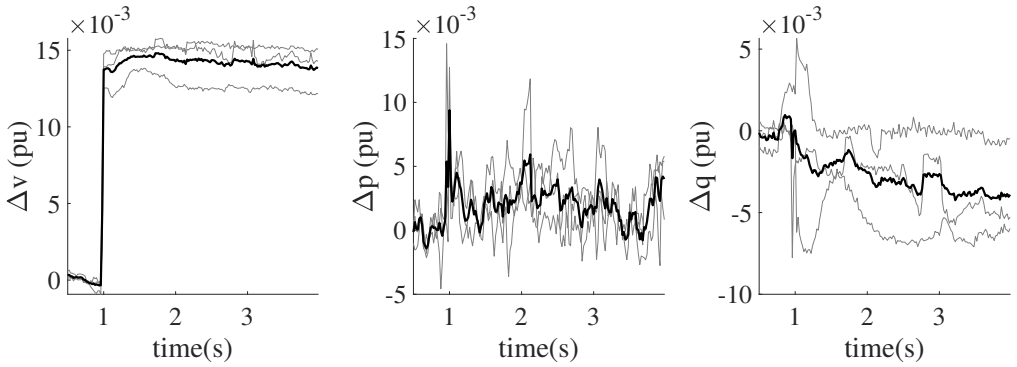


Figure 5.9: Sub.B.1: Indicative discarded cluster.

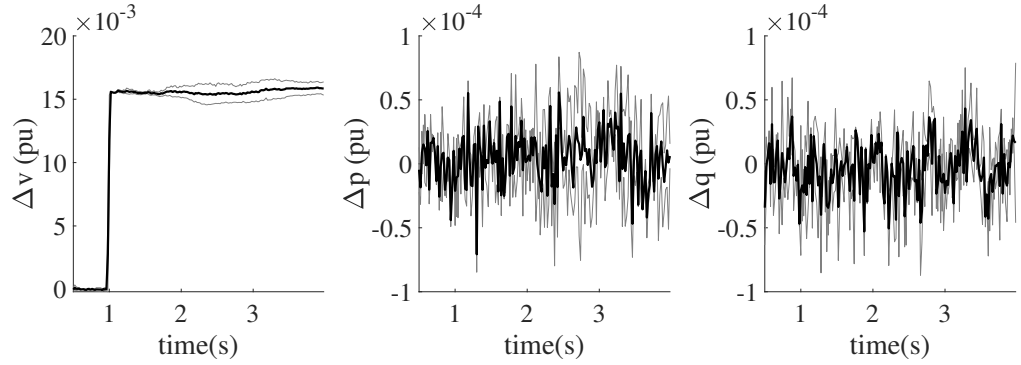


Figure 5.10: Sub.C.1: Indicative discarded cluster.

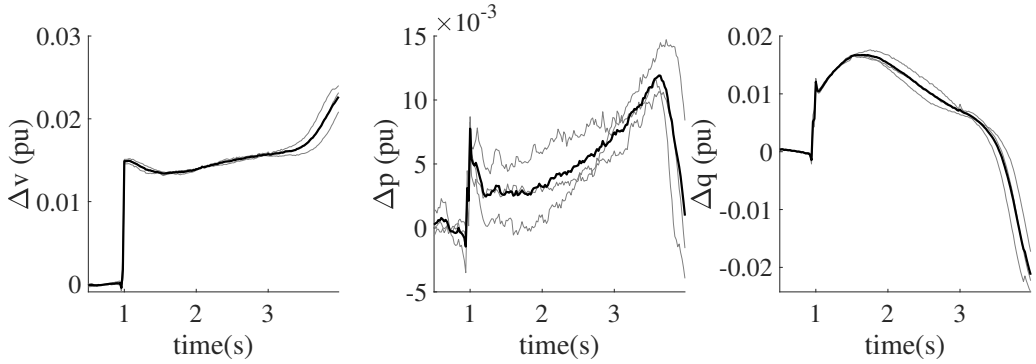


Figure 5.11: Sub.D.1: Indicative discarded cluster.

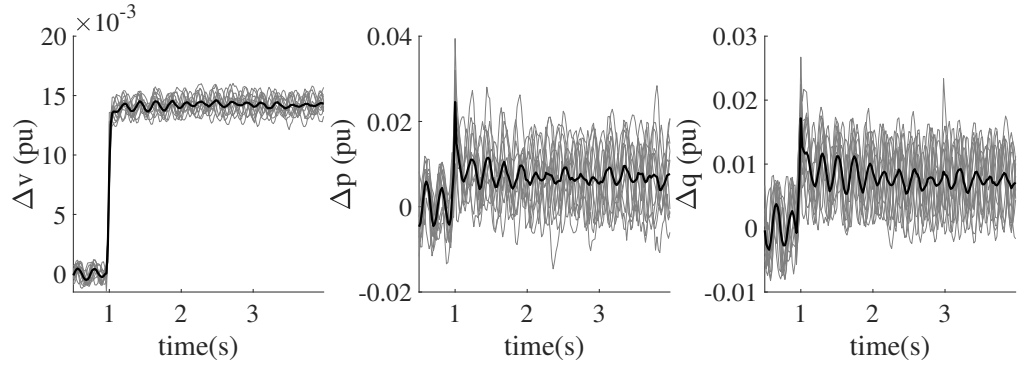


Figure 5.12: Sub.F.1: Indicative discarded cluster 1.

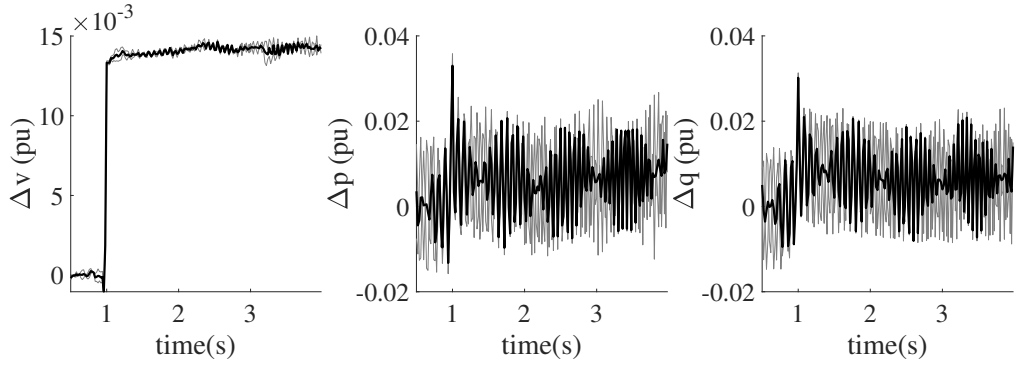


Figure 5.13: Sub.F.1: Indicative discarded cluster 2.

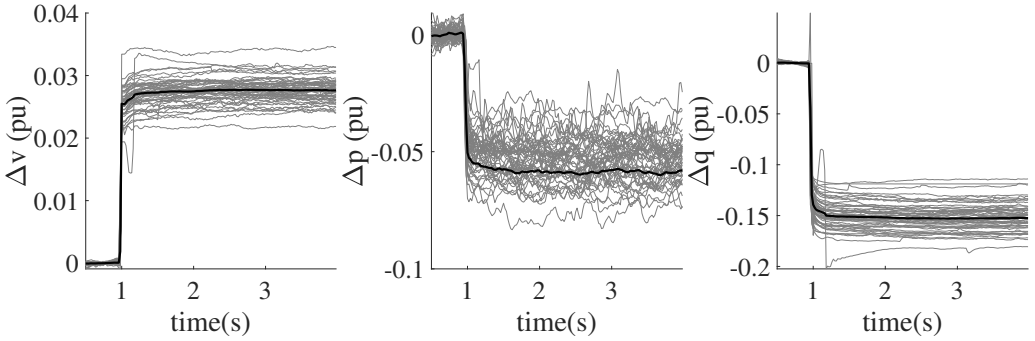


Figure 5.14: Sub.F.1: Indicative discarded cluster 3.

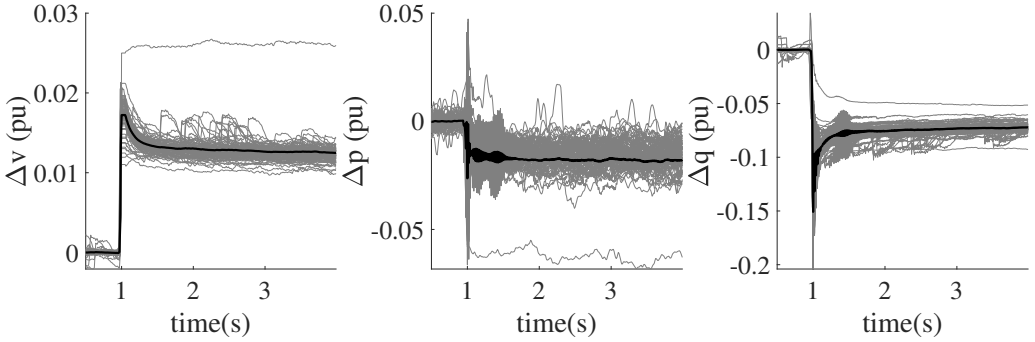


Figure 5.15: Sub.F.1: Indicative discarded cluster 4.

necessity of using the FD signals as features. Furthermore, Fig. 5.14 and 5.8 may describe the connection/disconnection of a big consumer within the examined ADN since both active and reactive power respond towards the opposite direction than the voltage. Similar type of response is found in Fig. 5.15, where a power controller of DG seems to have been activated. In addition, the clusters appeared in Fig. 5.9 and 5.10 contain highly noisy measurements that cannot be used for system identification. Finally, Fig. 5.11 illustrates a cluster with two measurements, which might describe a voltage stability issue. The reactive power starts increasing right after the disturbance and after around $1\,\mathrm{s}$, starts decreasing rapidly. To sum up, all those measurements should be omitted from the modeling procedure since they do not indicate the response (or a clear response) of an ADN after an external voltage disturbance. A possible use of those data may lead to faulty model parameters and error-prone models [128].

5.3.2 Stage 2

In stage 2, the user defines the optimal number of clusters using the identified dominant dynamics of stage 1. Thus, we can directly compare the algorithms' performance for the user-defined number of clusters. In this case study, one to three clusters of different dynamics were identified at each ADN. The deployed metrics are presented in Table 5.1-5.3. DBSCAN shows overall the best results, while k-means++ and agglomerative hierarchical follow with worse yet sufficient clustering results, as assessed by plotting separately the measurements of the derived clusters. Fuzzy c-means shows partly sufficient results, indicating that it could perform acceptably in a "cleaned" dataset without noise data and outliers. Note that the ADNs with no values in the tables comprise only one cluster of dynamics. Therefore, no clustering performance can be measured. To sum up, we conclude that DBSCAN is a suitable choice for clustering dynamic behavior using field measurements since it simultaneously discards outliers and groups the data accordingly without requiring a two-stage approach like the other algorithms. In addition, it runs without specifying the number of clusters *a priori*, which is considered a very challenging task in

Table 5.1: WCBCR for different ADNs and clustering algorithms.

ADN ID.	WCBCR			
	K-means++	Fuzzy c-means	Hierarchical	DBSCAN
Sub.A.1	6.2	6.2	6.2	6.22
Sub.A.2	8.2	11.18	9.64	7.3
Sub.B.1	18.26	18.26	19.09	13.51
Sub.B.2	11.48	3.21	11.82	3.95
Sub.C.1	180.93	184.13	197.57	160.92
Sub.C.2	-	-	-	-
Sub.D.1	2.26	65.56	2.46	2.61
Sub.D.2	-	-	-	-
Sub.E.1	24.59	27.78	29.42	24.35
Sub.E.2	-	-	-	-
Sub.F.1	149.47	147.05	138.12	98.32
Sub.F.2	9.84	48.9	9.85	6.45

cluster analysis [122].

As pointed out, the aim of this section is to highlight the existence of different dynamics within an ADN, which cannot be neglected during the modeling procedure. For each ADN, the "cleaned" dataset of voltage, active and reactive power measurements was grouped into distinguishable clusters. The clustering results of several representative ADNs are presented below. Similarly to the results of stage 1, the left graphs present the voltage change, the middle ones the active power response and the right ones the reactive power response. The grey curves indicate the individual recordings, whereas the black curves denote the average of each cluster's measurements. Moreover, comparative results between the average cluster curves are introduced.

Table 5.2: DBI for different ADNs and clustering algorithms.

ADN ID.	DBI			
	K-means++	Fuzzy c-means	Hierarchical	DBSCAN
Sub.A.1	2.46	2.46	2.46	2.48
Sub.A.2	2.93	4.27	3.5	2.45
Sub.B.1	4.27	4.27	4.35	3.58
Sub.B.2	2.61	2.44	2.56	2.49
Sub.C.1	13.43	13.56	13.32	12.1
Sub.C.2	-	-	-	-
Sub.D.1	9.4	11.69	9.04	4.66
Sub.D.2	-	-	-	-
Sub.E.1	3.29	3.47	37.75	3.24
Sub.E.2	-	-	-	-
Sub.F.1	36.05	22.75	35.69	14.28
Sub.F.2	27.43	33.3	27.48	2.38

Sub.A.1

At Sub.A.1, two clusters were generated as shown in Fig. 5.16-5.18. The major difference between those two clusters can be found in the reactive power response, whereas the active power responses are identical. In particular, the entries of cluster 2 are characterized by the normal recovery behavior of a distribution system without DG after a step voltage change. On the contrary, the entries of cluster 2 are characterized by the reactive power response depicted in 5.17. This phenomenon is caused by the voltage controller of the wind turbines installed within the examined ADN. This controller supports voltage through reactive current injections based on the corresponding ENTSO-E Network Code on Requirements for Generators [129].

Table 5.3: Silhouette index for different ADNs and clustering algorithms.

ADN ID.	Silhouette index			
	K-means++	Fuzzy c-means	Hierarchical	DBSCAN
Sub.A.1	0.98	0.98	0.98	0.98
Sub.A.2	0.79	0.73	0.78	0.81
Sub.B.1	0.87	0.87	0.86	0.9
Sub.B.2	0.86	0.89	0.86	0.88
Sub.C.1	0.81	0.8	0.79	0.81
Sub.C.2	-	-	-	-
Sub.D.1	0.62	0.58	0.65	0.83
Sub.D.2	-	-	-	-
Sub.E.1	0.78	0.72	0.76	0.78
Sub.E.2	-	-	-	-
Sub.F.1	0.56	0.56	0.57	0.84
Sub.F.2	0.59	0.48	0.59	0.85

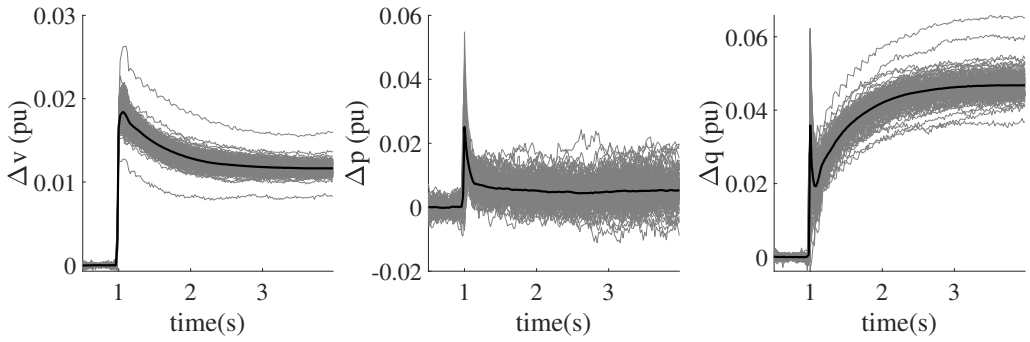


Figure 5.16: Sub.A.1: Cluster 1.

Sub.B.1

At Sub.B.1, two distinct clusters were generated as shown in Fig. 5.19-5.21. These two clusters are characterized by a similar active power response, in which the active

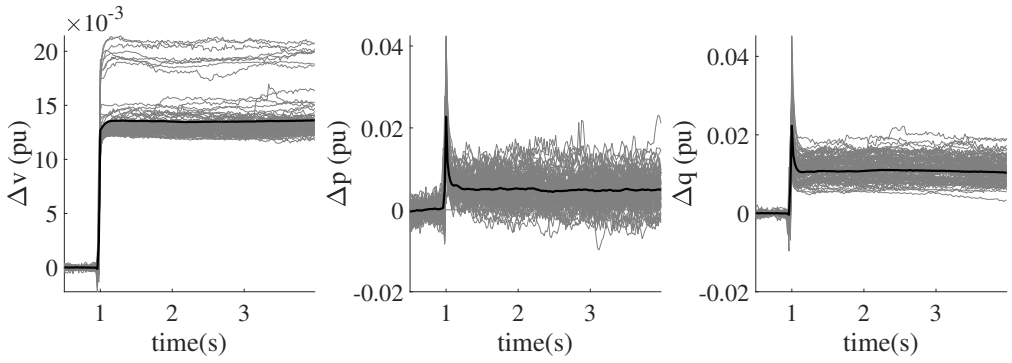


Figure 5.17: Sub.A.1: Cluster 2.

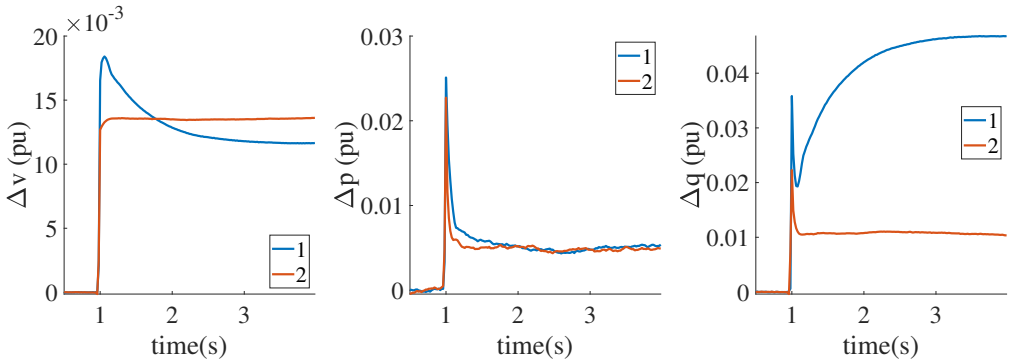


Figure 5.18: Sub.A.1: Average cluster curves.

power recovers partially to a new steady-state after a voltage change, as illustrated in the comparative plots of Fig. 5.21. In contrast, the reactive power response produced two distinguishable curves. Cluster 2 has got a relatively small reactive power change, in which the reactive power gradually recovers to a new steady-state, as depicted in Fig. 5.19. Cluster 1 is characterized by slower dynamics and their transient response is much higher than the one of cluster 2. This effect might be generated by the voltage controller of the DG units installed within the examined ADN.

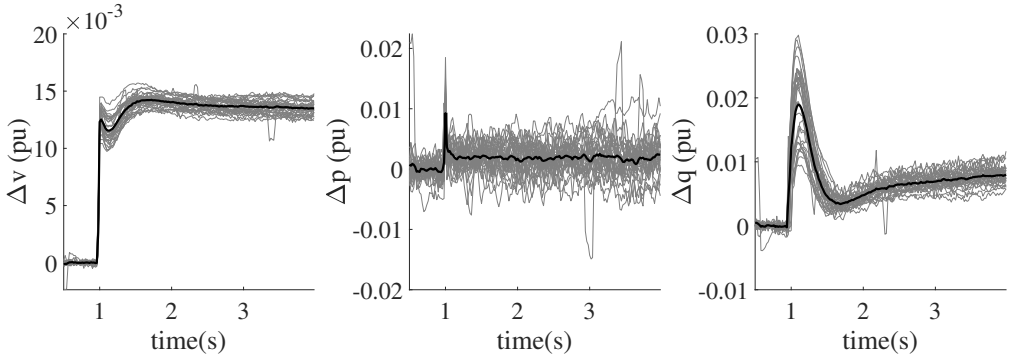


Figure 5.19: Sub.B.1: Cluster 1.

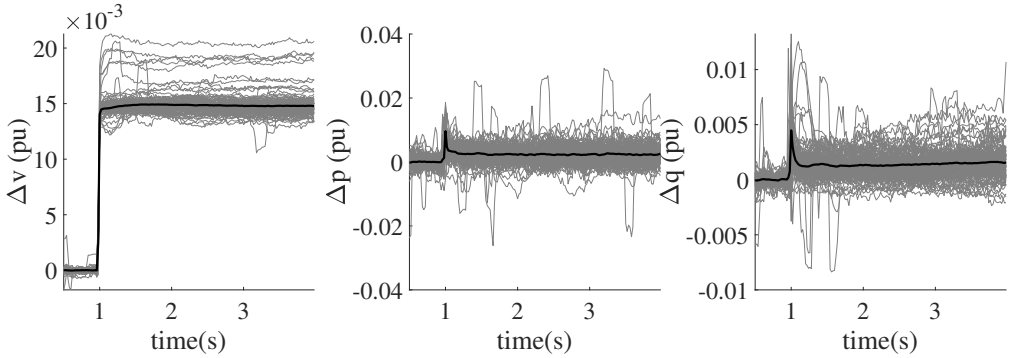


Figure 5.20: Sub.B.1: Cluster 2.

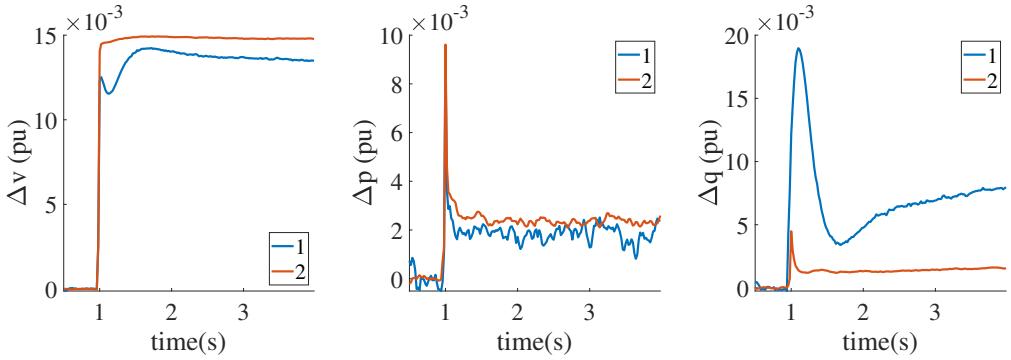


Figure 5.21: Sub.B.1: Average cluster curves.

Sub.C.1

At Sub.C.1 two different clusters were generated as depicted in Fig. 5.22-5.24. While both clusters present a similar recovery behavior in active power (with a small difference in the transient and steady state value), the reactive power response of cluster 2 is distinguished by slower dynamics as clearly observed in the comparative plots of Fig. 5.24. This particular reactive power response may be caused by the voltage controller of the CHP plant, which operates within the examined ADN and supports the voltage through reactive current injections.

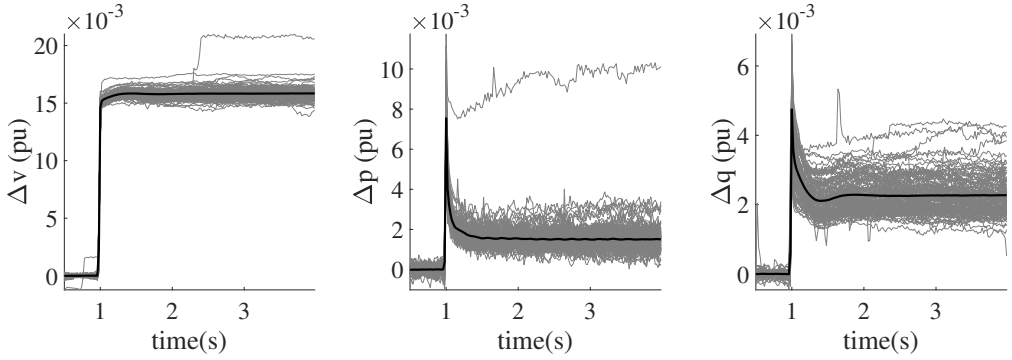


Figure 5.22: Sub.C.1: Cluster 1.

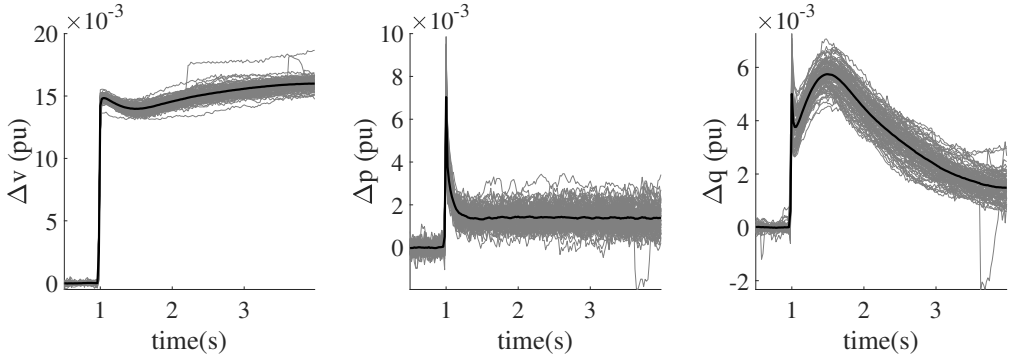


Figure 5.23: Sub.C.1: Cluster 2.

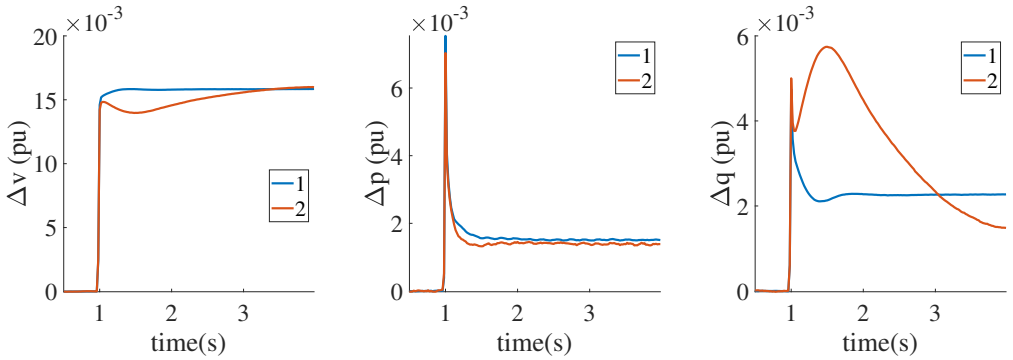


Figure 5.24: Sub.C.1: Average cluster curves.

Sub.D.1

At Sub.D.1 three different clusters were generated as shown in Fig. 5.25-5.28. All three clusters illustrate a common active power response without a significant difference between them. However, the reactive power response can highly vary among the three clusters as clearly depicted in the comparative plots of Fig. 5.28. Cluster 2 is characterized by a low frequency oscillation directly after the voltage disturbance. Importantly, this dynamic behavior may be caused by the two CHP plants, which are installed within the examined ADN and are equipped with voltage controllers. On the contrary, cluster 1 presents a different reactive power control, where the re-

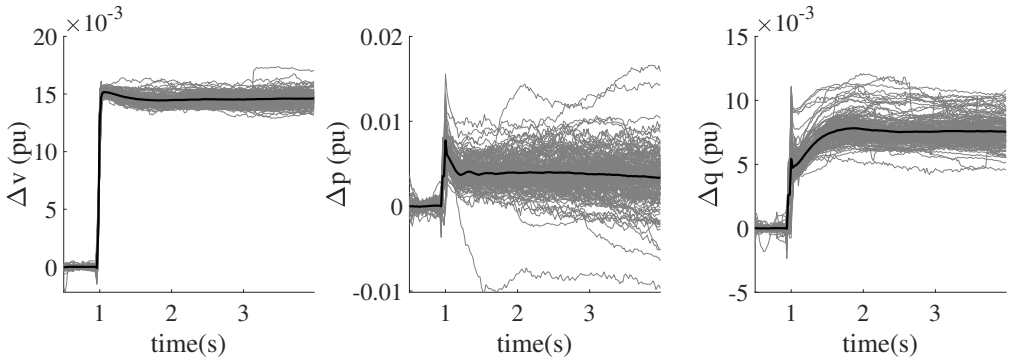


Figure 5.25: Sub.D.1: Cluster 1.

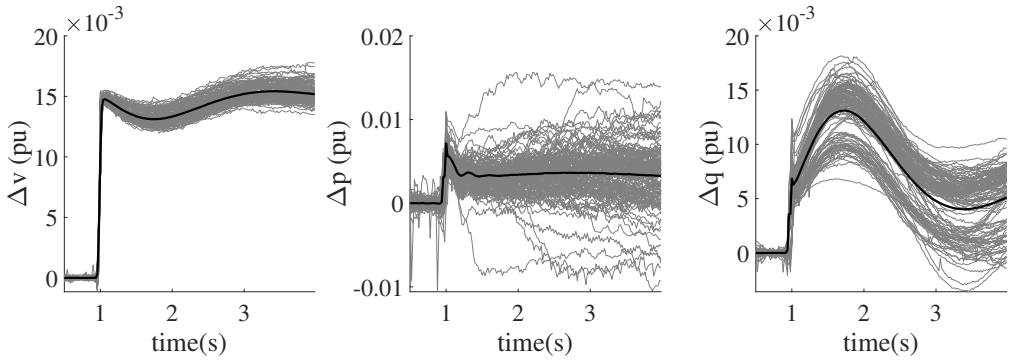


Figure 5.26: Sub.D.1: Cluster 2.

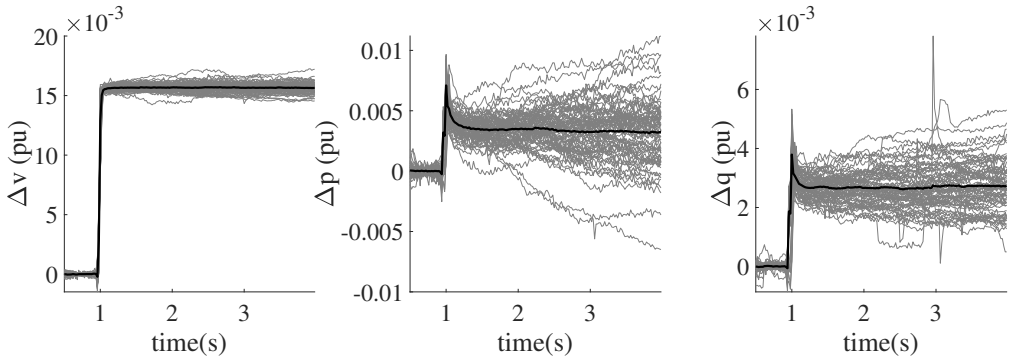


Figure 5.27: Sub.D.1: Cluster 3.

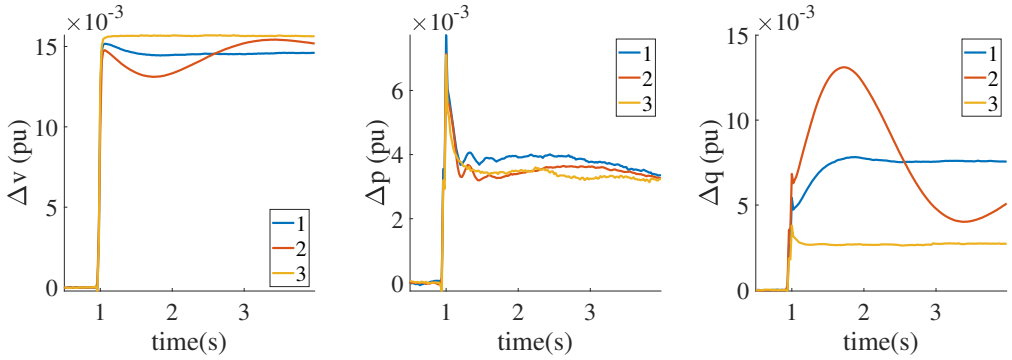


Figure 5.28: Sub.D.1: Average cluster curves.

active power converges to a remarkably higher value after the transient response. Finally, cluster 3 contains entries with the "normal" recovery part that have been observed in all ADNs under study.

Sub.E.2

At Sub.E.2, cluster 2 and 3 seem to have a similar active power recovery although they differ in the transient response, as shown in Fig. 5.32. On the other hand, cluster 1 is characterized by a big active power change and a small recovery as highlighted in the active power plots of Fig. 5.32. Regarding the reactive power, cluster 1 and 2 include measurements with a new steady-state quite close to the initial pre-disturbance value. Finally, cluster 3 shows a relatively moderate recovery compared to the other two clusters as illustrated in the reactive power plots of Fig. 5.32.

Sub.F.2

At Sub.F.2, three different clusters were generated as shown in Fig. 5.33-5.35. All three clusters show a dynamic response in which no voltage control is present. Their main differences can be observed in the transient and steady-state response. Cluster 1 and 2 appear to have similar dynamics with respect to active power but notably dif-

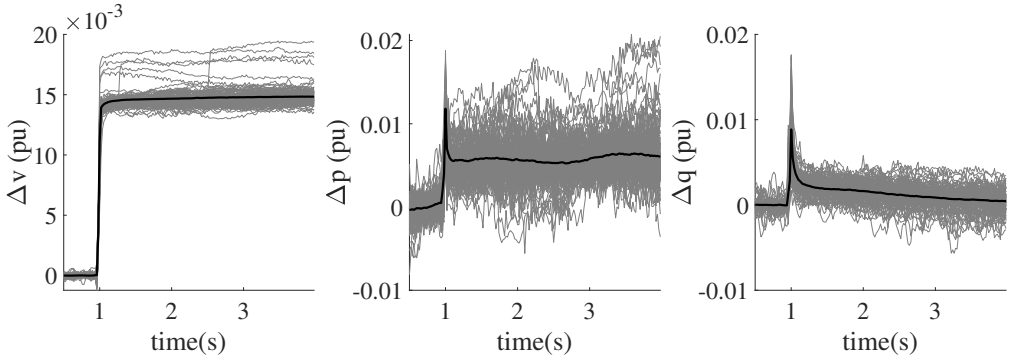


Figure 5.29: Sub.E.2: Cluster 1.

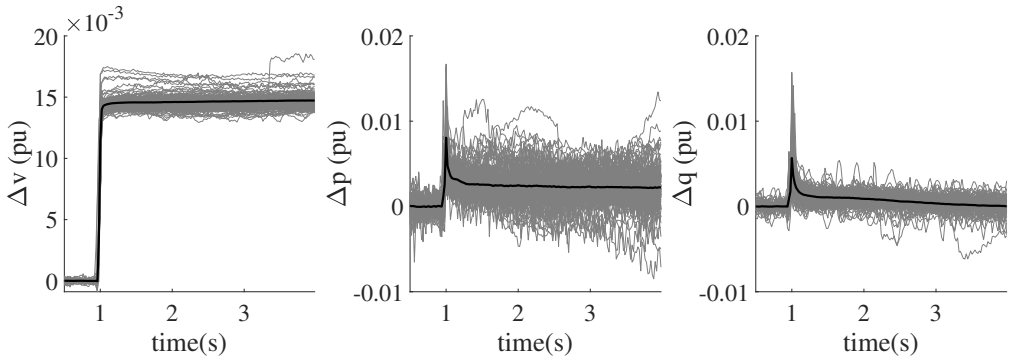


Figure 5.30: Sub.E.2: Cluster 2.

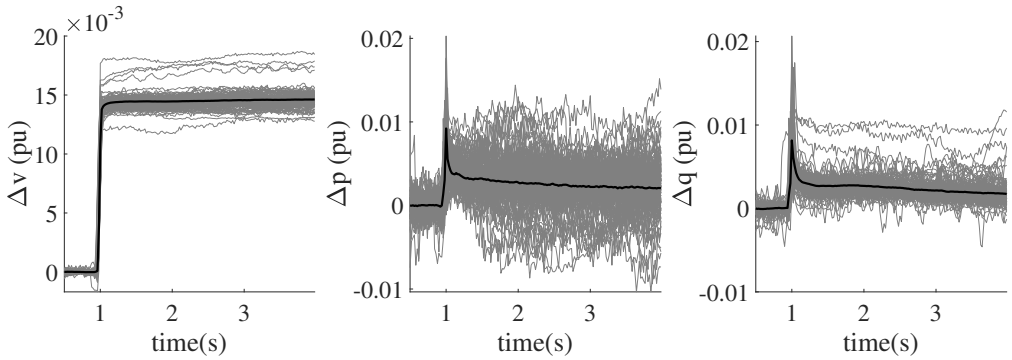


Figure 5.31: Sub.E.2: Cluster 3.

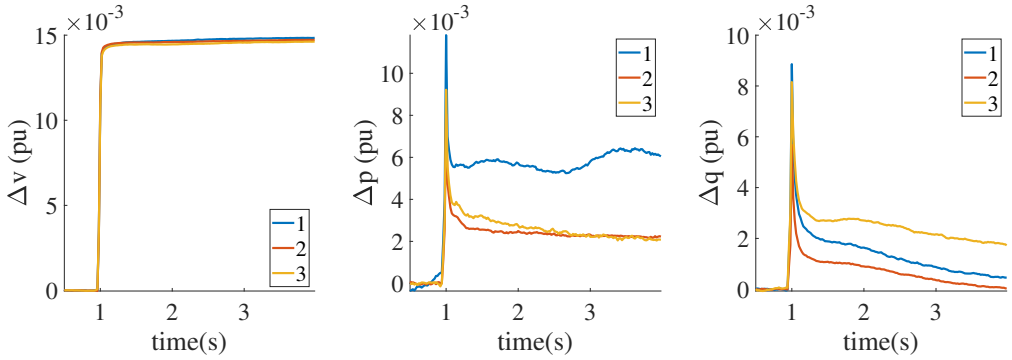


Figure 5.32: Sub.E.2: Average cluster curves.

ferent responses in reactive power, as depicted in the comparative plots of Fig. 5.36. Cluster 3 is characterized by a rather bigger active and reactive power change even though the voltage change is identical in all three clusters. This might be attributed to the different load composition across the various clusters.

Conclusion

As a general rule, the yielded clusters reveal similar dynamic responses in case of active power in all ADNs, where the load power demand recovers partially to a new steady-state after a voltage step, due to the presence of induction motors within the examined systems [67]. Regarding reactive power, several ADNs exhibit an entirely

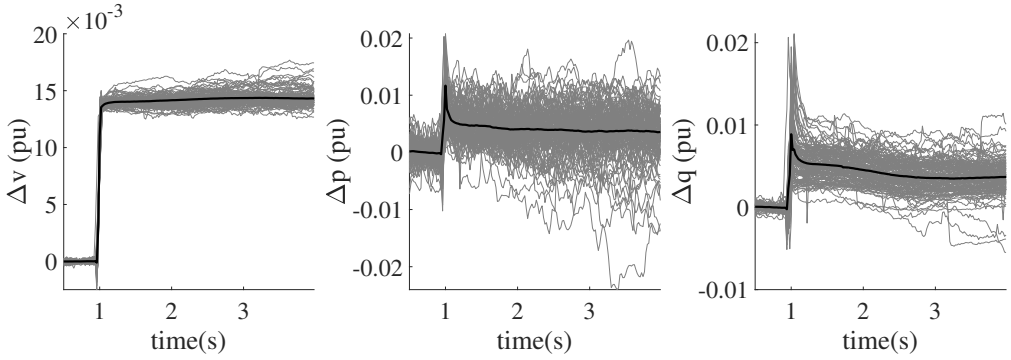


Figure 5.33: Sub.F.2: Cluster 1.

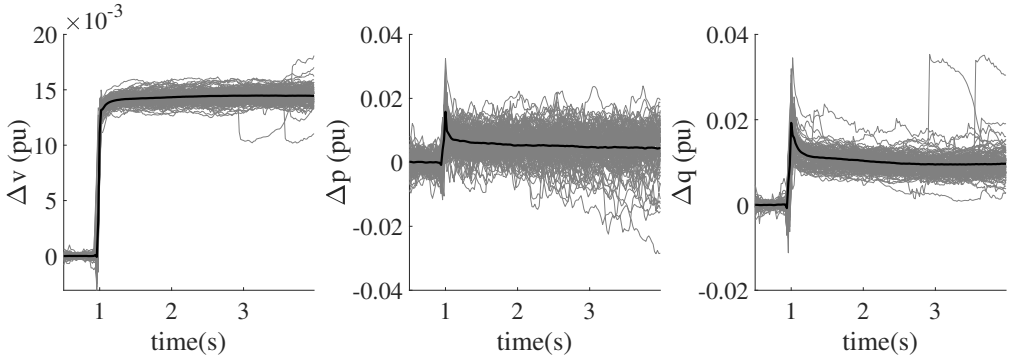


Figure 5.34: Sub.F.2: Cluster 2.

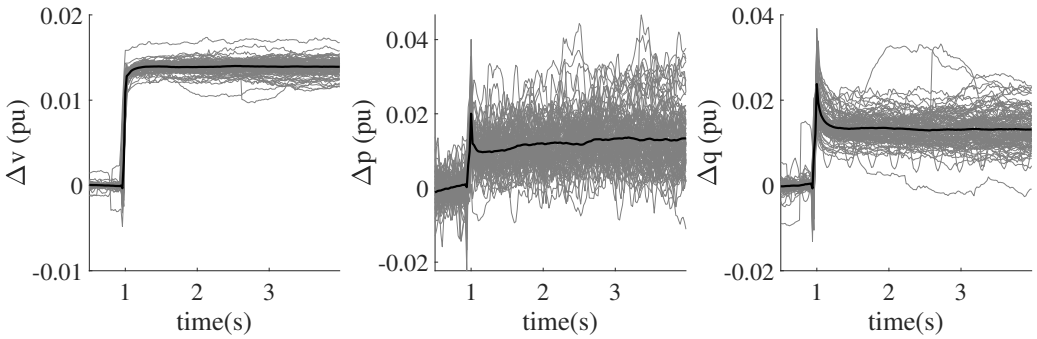


Figure 5.35: Sub.F.2: Cluster 3.

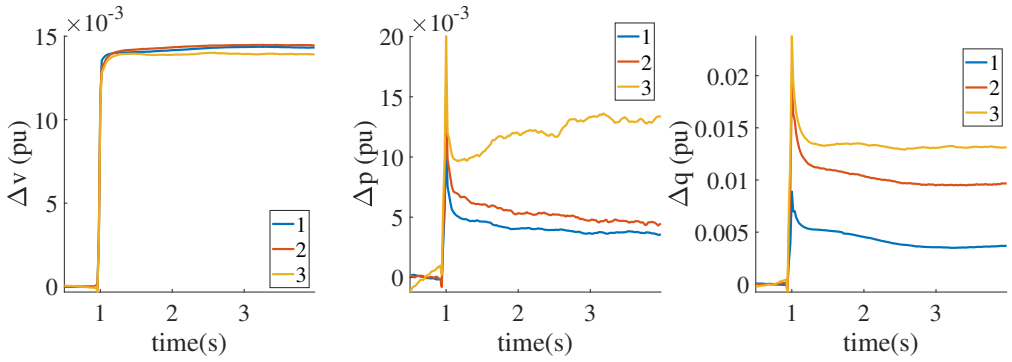


Figure 5.36: Sub.F.2: Average cluster curves.

different set of dynamics. Several ADNs with DG are characterized by one cluster with the classical load response to a voltage change like the one in Fig. 5.17 and at least one cluster with a dynamic voltage support scheme through reactive current injections based on the ENTSO-E Network Code [129], as in Fig. 5.16, 5.19, 5.23, 5.25, and 5.26. For example in Fig. 5.16, there is an immediate jump in reactive power due to the load reaction to the voltage change and after a few time steps, the DG units are activated and start supporting the voltage by injecting reactive power to the grid. Nevertheless, each control scheme is different and possibly highly dependent on the manufacturer of the DG unit.

5.4 Summary

Dynamic equivalent ADN models are commonly derived using the measurement-based approach. This method exploits acquired data in order to estimate the model parameters using system identification techniques. However, most of the approaches assume that the system maintains the same dynamics for different operating conditions, even though the load mix and the DG composition are constantly changing. At the same time, acquiring measurements over a significant period of time results in big datasets, where not all the recorded data are suitable for system identification.

To tackle those two major issues, an unsupervised learning approach is proposed comprising two stages. In the first stage, the irrelevant measurements are identified and are discarded from the modeling procedure. In the second stage, the remaining data are grouped into clusters of similar dynamic behavior. To do so, four clustering algorithms of different notion and complexity were examined. Overall, DBSCAN seems to be the best choice for this problem in practice. Furthermore, k-means++ and hierarchical clustering showed good performance and can be also deployed as alternative without jeopardizing the overall clustering performance.

Importantly, real field measurements acquired in the ADNs under study confirm the notable presence of irrelevant data within the recorded datasets due to the inability of the measurement units to classify the source of the disturbances. In the second stage, it was observed that an ADN may exhibit utterly different dynamics depending on whether the installed DG units are active or not. Therefore, unless an initial cluster analysis is performed, the direct deployment of even the "cleaned" dataset in the system identification procedure may lead to faulty models as demonstrated in the previous sections.

To sum up, the clustering results indicate that, indeed, an array of different dynamics can be present within ADNs due to diverse load and generation composition. This results justifies the imperative need of an initial cluster analysis or more generally, the need for a preprocessing technique. It is worth pointing out that no knowledge of load and generation mix is required and the clustering is performed exclusively based on the measurements themselves. Moreover, the proposed method can

be easily integrated as a preprocessing step into most of the existing ADN modeling approaches in order to obtain an insight of the governing dynamics characterizing the examined systems.

Chapter 6

Linear Dynamic ADN Model

This chapter presents a linear dynamic model that can be easily integrated to any power system simulation software available in the market. It starts with the parameter estimation of a classical active power model and then, introduces a new approach for modeling the reactive power. This approach is based on separating a reactive power response into two different signals. The first one is attributed to the load and the second one is attributed to the generation. In doing so, two different models, i.e., one for load and one for generation, can be derived based on those two signal categories. This chapter ends with a comparison between the dynamic load model parameters estimated several years ago and now.

6.1 Active power modeling

The aim of this section is to develop per unit (pu) models, which can be easily scaled based on the corresponding voltage level and the real power consumption. As stressed out throughout this thesis, the power recorded at the PCC does not correspond to the real load demand, since there is DG installed within the examined grids. Therefore, in order to obtain the actual power consumption, the measured power at PCC should be calibrated using the corresponding DG values. In this regard, dynamic active power models can be developed only for the substations

with complete data.

However, the exact DG information at the moment of the disturbance was not available. Instead, the DSOs provided 15-minute resolution time series for some of the big DG plants within the examined grids, where each point of the given data represents the average power generation of the previous 15 minutes. In this context, the given data are interpolated accordingly in order to estimate the power generation at the moment of the voltage disturbance. Then, this result is used to estimate the actual load consumption.

In this study, the available DG data allow an accurate estimation of the actual load in the following ADNs:

- Sub.A.2
- Sub.B.1
- Sub.C.1
- Sub.C.2
- Sub.D.1
- Sub.D.2
- Sub.E.1

6.1.1 Preprocessing

In this section, a simple procedure is presented in order to estimate the pu input (voltage) and output signals (active power). As a first step, the actual active power $\hat{P}_i(t)$ is estimated using:

$$\hat{P}_i(t) = P_{\mathrm{meas},i}(t) + \hat{P}_{\mathrm{DG},i} \tag{6.1}$$

where $P_{\mathrm{meas},i}(t)$ is the i-th active power response as recorded by the measurement unit and $\hat{P}_{\mathrm{DG},i}$ is an estimate of the total DG at the moment of the disturbance based

on the linear interpolation of the 15-minute time series. It is assumed that $\hat{P}_{\mathrm{DG},i}$ is constant over the time period of the corresponding measurement.

In this context, the initial actual load demand before the voltage disturbance is estimated as:

$$\hat{P}_{0,i} = \frac{1}{N_0} \cdot \sum_{t_n \in [t_0 - 1\,\mathrm{s}, t_0)} \hat{P}_i(t_n) \tag{6.2}$$

where t_0 is the time of the disturbance and N_0 is the number of samples recorded within $1\,\mathrm{s}$ before the disturbance. At this point, all the measurements that have $\hat{P}_{0,i} < 1\,\mathrm{MW}$ are discarded from the modeling procedure, since $1\,\mathrm{MW}$ is considered too low to be the real load demand of a distribution system and those power values may be caused by missing or incorrect DG data. Furthermore, highly noisy measurements are also discarded using the methodology described in [39]. This methodology classifies a measurement as highly noisy if the signal variation during the new steady-state is bigger than the 1/3 of the transient response.

Finally, we can estimate the pu active power change Δp_i for every measurement i as

$$\Delta p_i(t) = \frac{\hat{P}_i(t) - \hat{P}_{0,i}}{\hat{P}_{0,i}}. \tag{6.3}$$

As for the voltage, the procedure to calculate the corresponding voltage signals is straightforward since the measured voltage $V_i(t)$ is the actual voltage. Thus, the pu change Δv_i for each measurement i can be calculated using

$$\Delta v_i(t) = \frac{V_i(t) - V_{0,i}}{V_{0,i}}, \tag{6.4}$$

where $V_{0,i}$ is the average voltage level $1\,\mathrm{s}$ before the disturbance and is estimated, similarly to the case of active power, by

$$V_{0,i} = \frac{1}{N_0} \cdot \sum_{t_n \in [t_0 - 1, t_0)} V_i(t_n). \tag{6.5}$$

6.1.2 Parameter estimation

After estimating the pu signals for voltage and active power, a dynamic model is estimated for each of the aforementioned ADNs. A simplified schematic diagram of

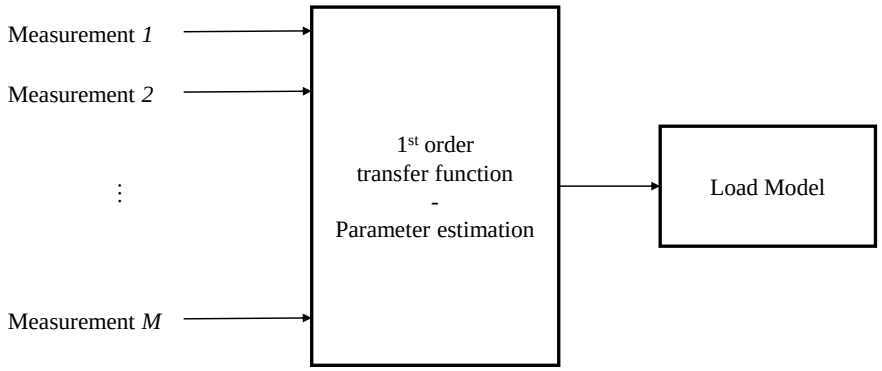


Figure 6.1: System identification for load models.

this procedure is shown in Fig. 6.1, where each measurement corresponds to the pair ($\Delta v_i(t)$, $\Delta p_i(t)$). Importantly, a first order transfer function was selected to model the active power response of the system based on the inspiring work presented in [130]. Finally, the parameters of each transfer function are estimated using the CONtinuous-Time System IDentification (CONTSID) technique [131].

6.1.3 System identification results

In this section, we compare the estimated dynamic models of each ADN. To do so, we apply a common input signal to every transfer function and then, we plot the generated dynamic responses in a common graph. The input signal is a voltage step change of 0.02 pu, as shown in Fig. 6.2. The generated active power responses are illustrated in Fig. 6.3. Note that the response with the label "Hall" refers to the model that was developed in [130] several years ago. This old model is employed for two reasons. The first one is to observe how the load dynamics have changed throughout the years. The second one lies in the fact that the work of [130] is one of the very limited studies using real field data acquired in several distribution networks.

As depicted in Fig. 6.3, the transient responses range from 0.08 to almost 0.15

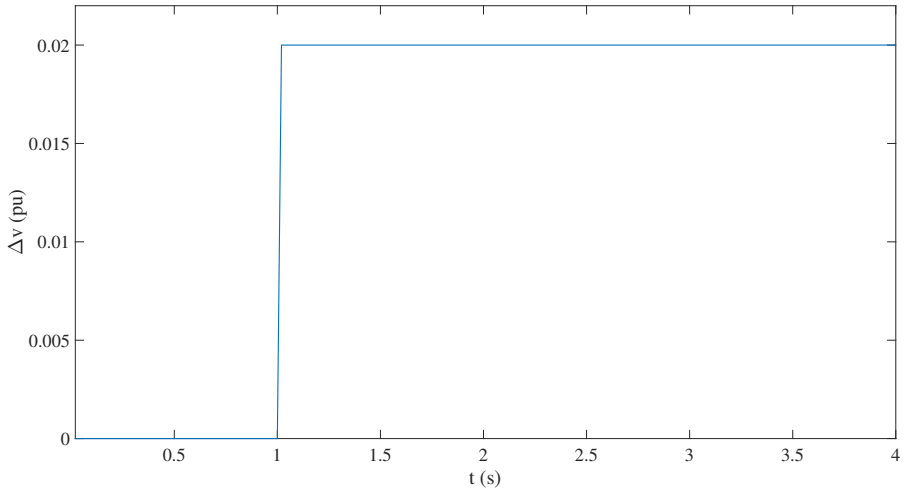


Figure 6.2: Voltage step change of 0.02 pu.

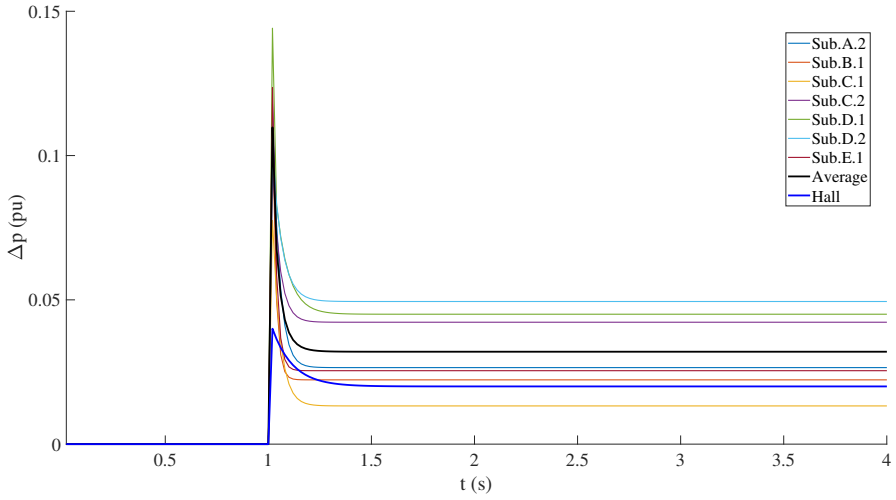


Figure 6.3: Active power response of each ADN after a voltage step change of 0.02 pu.

pu whereas the steady-state responses vary from 0.02 to 0.05 pu. Importantly, all the ADNs recover to a new steady-state around $0.2\,\mathrm{s}$ after the voltage disturbance. To compare those results with the results obtained in the past as presented in [130]

(bold blue curve in Fig. 6.3), we estimate an average active power response using all the available responses of the examined ADNs. The resulted average response is shown in Fig. 6.3 as the bold black curve. As observed in this graph, the model from [130] has a significant smaller transient jump and recovers slower to the new steady-state. On the other hand, it seems that there is no big difference between the two steady-state gains. It is worth pointing out that the difference in the transient response between the two models might be caused by the measurement equipment used around 30 years ago, which do not provide the same high sampling frequency as the new modern devices.

6.1.4 Generic model

In order to obtain an "average" model, which can be deployed in dynamic simulations, the parameters of a transfer function

$$G_{pv}(s) = k_{pv}\frac{1+sT_2}{1+sT_1} \tag{6.6}$$

are estimated using all the available responses. The estimated k_{pv} of the average model corresponds to the steady-state gain of the model and is also presented in Table 6.1 alongside with the other k_{pv}. Here, it is observed that the average steady-state gain estimated in this study is $k_{pv} = 1.60\,\%/\%$, which is higher than the value of 1.0 %/% observed in the past. However, it can be noted that the latter lies within the range of values observed in the derived models of this work. Finally, the transfer function of this average model can be written as

$$G_{pv}(s) = 1.6033\frac{1+0.0985s}{1+0.0291s}, \tag{6.7}$$

and can be readily integrated into any commercial software for power systems simulation as a realistic load model for electromechanical transients.

6.2 Reactive power modeling

The aim of this section is to introduce the procedure for developing linear dynamic models for reactive power. As discussed in Chapter 3, the reactive power response

Table 6.1: k_{pv} of the examined ADNs.

ADN ID	k_{pv} (%/%)
Sub.A.2	1.33
Sub.B.1	1.11
Sub.C.1	0.66
Sub.C.2	2.11
Sub.D.1	2.25
Sub.D.2	2.47
Sub.E.1	1.27
Average	1.60
Average in [130]	1.0

significantly depends on whether and which DG unit is present and active. To take this into account, the main idea is to create two different dynamic models: one for the reactive power response of the loads and one for the reactive power response of the DG. Then, the sum of those two individual responses will be the reactive power response of the system.

To do so, instead of using the whole dataset of measurements for system identification as in case of active power, the whole dataset will be divided into subsets with and without reactive power control. The main reason for splitting the dataset based on this criterion lies in the fact that the DG systems are operating using this kind of control whereas load components are not.

As in case of active power, the reactive power response of the loads will be modeled in per unit (pu) in order to be scaled easily for various voltage and reactive power levels. The main issue that arises here is that there is no information about the amount of reactive power that the various DG systems consume or inject to the distribution system. Therefore, the measured reactive power at the PCC is essentially composed by the reactive power of the loads, possibly power factor correction devices, and the reactive power of DG. Note that a fully compensated distribution

network, i.e. $Q_0 = 0$ does not yield $\Delta Q = 0$ for voltage changes and that $Q_0 = 0$ is not a suitable reference value. However, it is plausible that the change in reactive power will be proportional to the size of the load, which is known by $\hat{P}_0$. To this end, it is assumed that the active power loads are operating with a constant power factor in order to be able to model the reactive power in pu. In particular, a constant power factor of loads (0.78) without power factor correction such that $\tan\varphi = 0.8$ is assumed. Note that this factor of 0.8 is somewhat arbitrary and that another factor (such as 1) would work as well, as long as the same value is used for the system identification and in the power system model in which the model is embedded. Hence, the reactive power $\hat{Q}_{i,\mathrm{ref}}$ used for the conversion to pu values can be written as

$$\hat{Q}_{i,\mathrm{ref}} = 0.8 \cdot \hat{P}_{0,i}, \tag{6.8}$$

where $\hat{P}_{0,i}$ is an estimate of the actual load demand as calculated in the previous chapter for the following distribution grids:

- Sub.A.2
- Sub.B.1
- Sub.C.1
- Sub.C.2
- Sub.D.1
- Sub.D.2
- Sub.E.1

However, the reactive power responses of DG are formed by the control strategy of the individual DG systems. As a result, these responses neither are proportional to the actual power of the DG units, nor to the power of the loads. Rather, the reactive power responses scale with the installed capacity of DG that is equipped with corresponding control functionality. For this reason, using as base the values

of $\hat{P}_{\mathrm{DG},0}$ or even $\hat{P}_0$ might lead to error-prone models. Therefore, the reactive power response of DG is modeled using absolute values. Note that this is indicated by the use of a capital letter ΔQ_{DG} for the output of this model in Fig. 6.4. It is worth pointing out that in order to use these models in a simulation platform, the input voltage and the output reactive power must be scaled according to the respective values.

To sum up, Fig. 6.4 describes the overall model structure that is employed for modeling both active and reactive power.

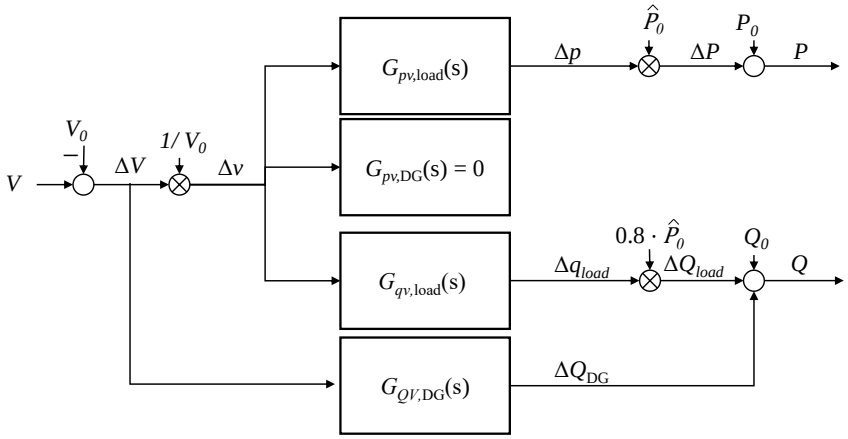


Figure 6.4: Block diagram representation of the model structure for active and reactive power. In case of distribution grids in which no DG with reactive power control is present, $G_{QV,\mathrm{DG}}(s) = 0$.

As there are only a few ADNs for which a significant influence of DG units on the reactive power response was observed, the proposed methodology is applied only on the measurement data from the following ADNs:

- Sub.A.1
- Sub.B.1
- Sub.C.1

- Sub.D.1

For all other ADNs, no DG impact was observed on the reactive power dynamics. In those cases, the corresponding DG units do not provide voltage control. For these ADNs, the classical load modeling approach is applied using the whole dataset, analogously to the procedure for active power described in previous section.

6.2.1 Methodology

In this section, the methodology deployed for generating the reactive power models for loads, $G_{qv,\mathrm{load}}(s)$, and for DG, $G_{QV,\mathrm{DG}}(s)$ is described. The proposed methodology follows the flowchart depicted in Fig. 6.5.

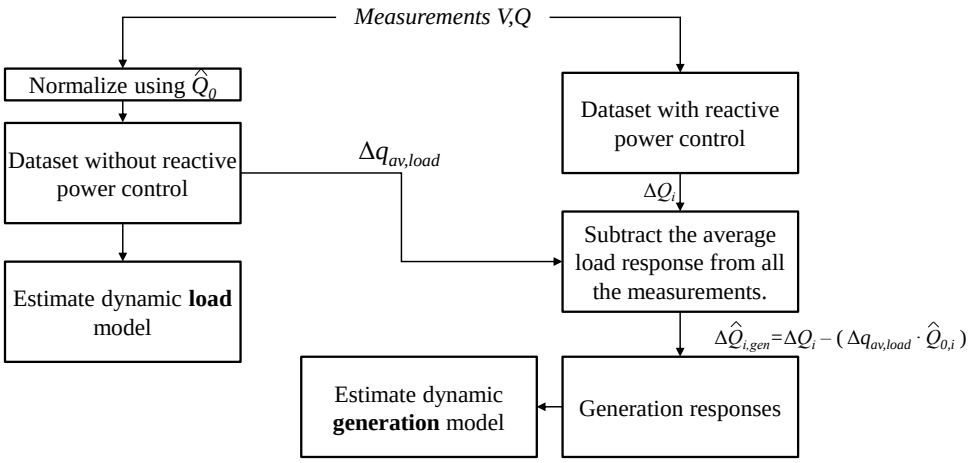


Figure 6.5: Flowchart illustrating the process of building linear reactive power models.

As a first step, it is necessary to decide whether voltage controller dynamics can be observed in the measurement data. If yes, the measurement dataset is to be split into subsets. A number of two subsets has shown to be sufficient for the ADNs under consideration in this study. If an ADN contains several DG of sufficient size that provide voltage control in a different way, a larger number of subsets may be necessary.

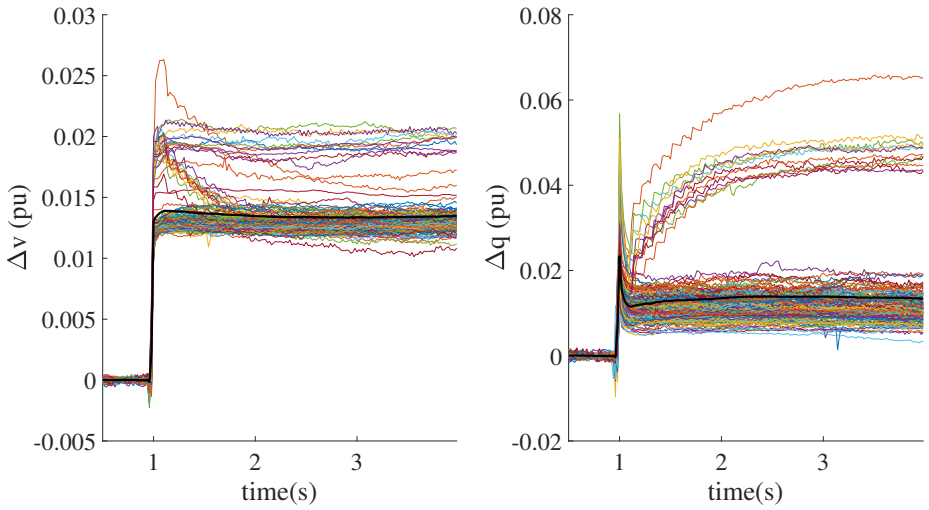
Initially, the split was performed based on the given DG data, and especially based on whether the DG units were generating power. However, this approach resulted in subsets in which many, but not all of the measurements of each group exhibit similar dynamic behavior. Fig. 6.6 shows an example of the resulting subsets when the dataset is split based only on the information of the given DG data. Both graphs (upper and lower) present the pu change in voltage and reactive power for Sub.A.1. The lower graph (6.6b) contains the measurements in which a wind park was generating power (according to the DG time series data). The upper graph (Fig. 6.6a) contains all the other measurements, i.e., those for which no generation from wind parks was recorded. As observed in the upper graph, there are entries that follow the classical load response as well as entries that contain voltage controller dynamics. Similarly, the lower plots also contain both types of responses. Hence, it is clear that splitting the measurements using only the DG information does not generate homogeneous groups. To tackle this problem, the proposed unsupervised learning method for grouping similar dynamic responses was deployed as presented in Chapter 5.

Based on the clustering results of the proposed unsupervised learning method, the new generated groups using this clustering approach are shown in Fig. 6.7. The clustering algorithm successfully divided the data into two homogeneous groups. In the following section, the measurements of the first group (Fig. 6.7a) are used to build a reactive power model for pure loads, while the measurements of the second group (Fig. 6.7b) are combined with the pure load model to obtain a reactive power model for DG.

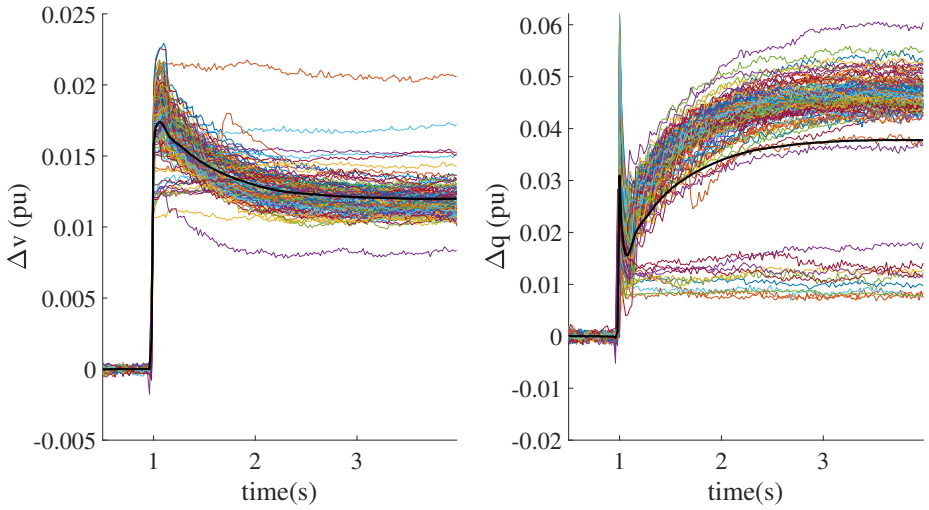
6.2.2 Reactive power model for loads

This subsection corresponds to the left flow of Fig. 6.5. Once the measurements have been separated into two (or more) groups, the measurements that correspond to the load responses are normalized using the equation:

$$\Delta q_i(t) = \frac{Q_i(t) - \hat{Q}_{0,i}}{\hat{Q}_{i,\text{ref}}} \tag{6.9}$$



(a) Measurements with zero DG.

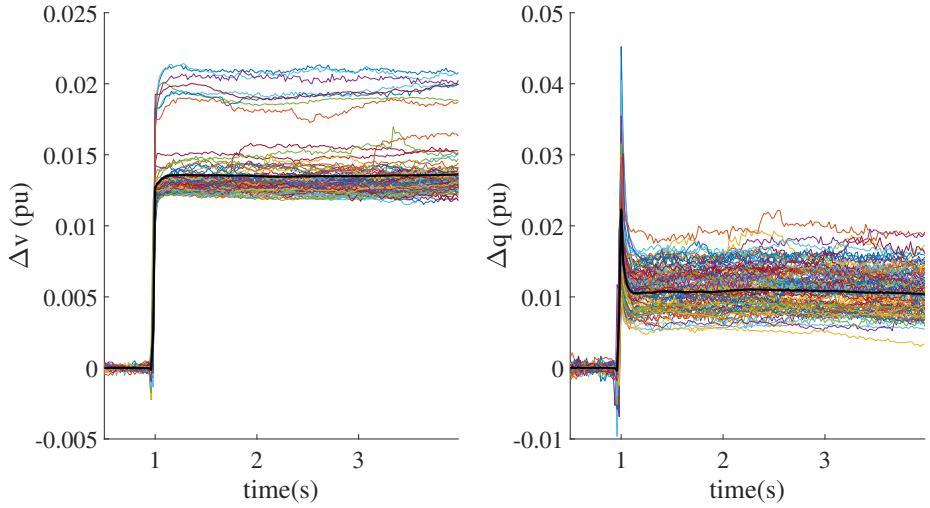


(b) Measurements with DG generating power.

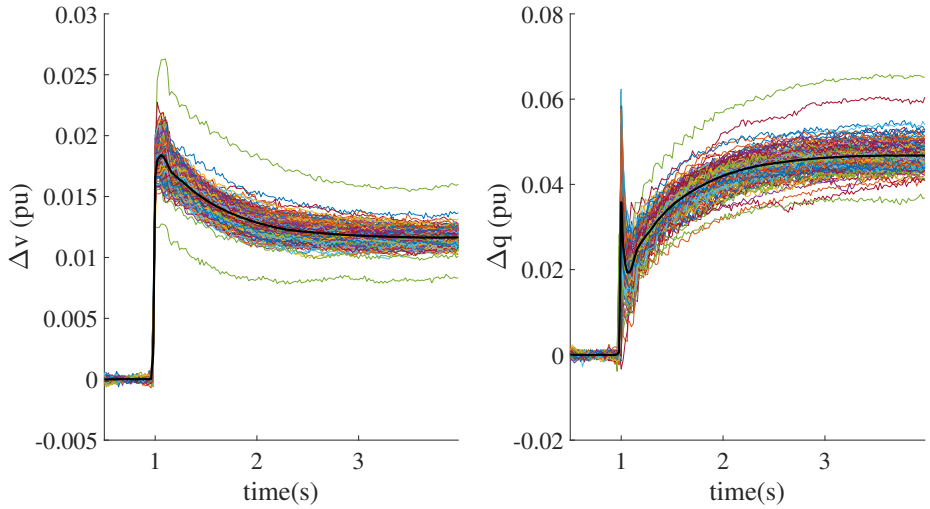
Figure 6.6: Split based only on the given DG data.

where $\Delta q_i(t)$ is the pu response of reactive power for the i-th measurement, $Q_i(t)$ is the recorded reactive power, $\hat{Q}_{i,\text{ref}}$ is defined by (6.8), and the initial reactive power before the voltage disturbance is estimated as:

$$\hat{Q}_{0,i} = \frac{1}{N_0} \cdot \sum_{t_n \in [t_0 - 1\,\text{s}, t_0)} \hat{Q}_i(t_n). \tag{6.10}$$



(a) Cluster with the classical load recovery behavior.



(b) Cluster with voltage controller dynamics.

Figure 6.7: Split based on the proposed clustering algorithm.

The pu voltage signals are estimated using the equations (6.4) and (6.5). After estimating the pu signals for voltage, $\Delta v_i(t)$ and reactive power, $\Delta q_i(t)$, a dynamic model is estimated for each of the available ADNs listed in the beginning of this chapter. A simplified schematic diagram of this procedure is shown in Fig. 6.1, where each measurement corresponds to the pair ($\Delta v_i(t)$, $\Delta q_i(t)$). Importantly, a

first order transfer function was selected to model the reactive power response of the system based on the inspiring work presented in [130]. Finally, the parameters of each transfer function are estimated using the CONtinuous-Time System IDentification (CONTSID) technique [131].

System identification results

In this section, the estimated dynamic models of each distribution network are compared. To do so, a common input signal is applied to every transfer function and then, the generated dynamic responses are plotted in a common graph. The input signal is a voltage step change of 0.02 pu as shown in Fig. 6.8. In this context, the generated reactive power responses are illustrated in Fig. 6.9.

As depicted in Fig. 6.9, the transient responses range from around 0.05 to almost 0.13 pu whereas the steady-state responses vary from 0.02 to 0.05 pu. Importantly, all the distribution networks recover to a new steady-state around $0.2\,\mathrm{s}$ after the voltage disturbance. To compare those results with the results that were obtained in the '80s/'90s and are presented in [130] (bold blue curve in Fig. 6.9), an average reactive power response is estimated using all the available responses of the exam-

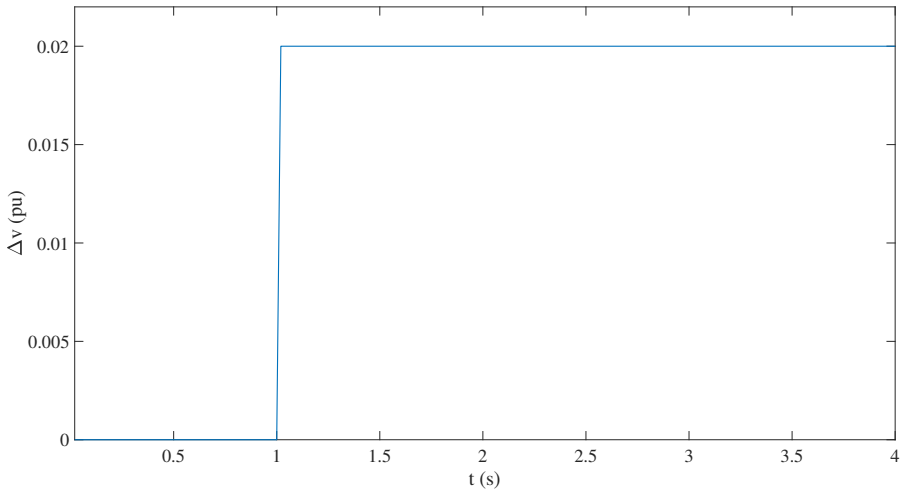


Figure 6.8: Voltage step change of 0.02 pu.

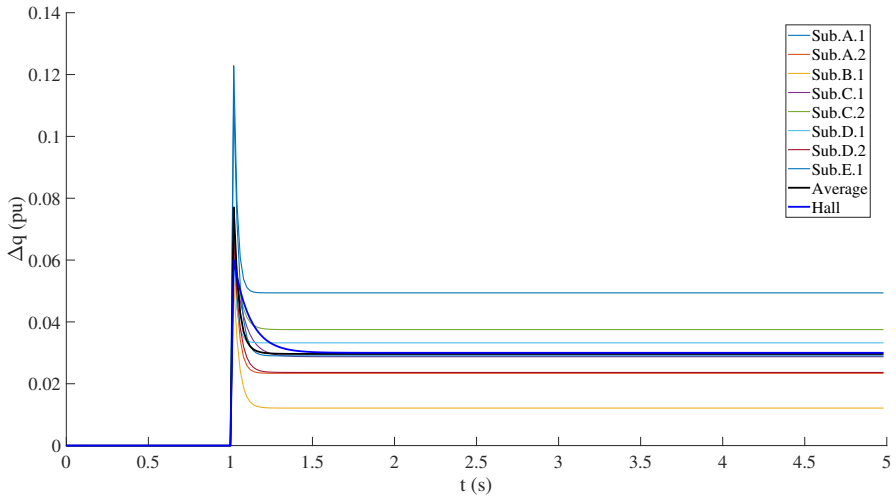


Figure 6.9: Average reactive power response of each ADN after a voltage step change of 0.02 pu.

ined ADNs. The resulting average response is shown in Fig. 6.9 with the black bold curve. The old model has a slightly smaller transient jump and recovers slower in the new steady-state. On the other hand, there is almost no difference between the two DC gains. To better examine the DC gain, we assume a transfer function

$$G_{qv,load}(s) = k_{qv}\frac{1+sT_2}{1+sT_1}. \tag{6.11}$$

The estimated k_{qv} are presented in Table 6.2. Here, it is observed that the average DC gain has remained the same within the past years, maintaining a value of 1.5 %/%.

Generic model

In this section, a generic model, which can be deployed in dynamic simulations, is estimated using all the available reactive power responses of the examined ADNs. The derived transfer function can be written as

$$G_{qv}(s) = 1.4822\frac{1+0.0821s}{1+0.0317s}, \tag{6.12}$$

Table 6.2: k_{qv} of the examined distribution networks.

ADN ID	k_{qv} (%/%)
Sub.A.1	2.47
Sub.A.2	1.17
Sub.B.1	0.60
Sub.E.1	1.45
Sub.C.1	1.44
Sub.C.2	1.88
Sub.D.1	1.66
Sub.D.2	1.18
Sub.E.1	1.45
Average	1.48
Average in [130]	1.5

and can be readily integrated into any commercial software for power systems simulation as a realistic load model.

6.2.3 Reactive power model for generators

In this section, the various steps of the right flow of Fig. 6.5 are introduced. The first step is to generate the dataset that includes only the measurements with voltage control. This step was done using the clustering algorithm that was described in Chapter 5. The second step is to subtract the average load response of the first group, denoted by the subscript *load* from each individual measurement of the second group, denoted by the subscript *DG*. This step attempts to isolate the dynamic responses caused by the DG units by removing an estimate of the load response. To do so, for every measurement i of the second group, we estimate the reactive power response for generators $\Delta\hat{Q}_{i,\text{gen}}$ as

$$\Delta\hat{Q}_{i,\text{gen}} = \Delta Q_i - \left(\Delta q_{\text{av,load}} \cdot \hat{Q}_{i,\text{ref}}\right) \tag{6.13}$$

where $\Delta Q_i = Q_i(t) - \hat{Q}_{0,i}$ and

$$\Delta q_{\text{av,load}} = \frac{1}{N_{\text{load}}} \cdot \sum_{i=1}^{N_{\text{load}}} \Delta q_{i,\text{load}}(t), \tag{6.14}$$

with N_{load} being the number of measurements contained in the load subset.

The resulting $\Delta\hat{Q}_{i,\text{gen}}$ for the case of Sub.A.1 are shown in Fig. 6.10 in pu. The respective plots for the other ADNs are shown in Appendix. However, the derived responses show some spikes at the moment of the disturbance, which are caused by the difference between the i-th load response and the average load response. In general, at the moment of voltage step, the induction motors' slip cannot change and thus, the aggregate load behaves as static [67]. Thus, the magnitude of the reactive power jump is exclusively determined by the load amount and composition (at that particular moment). Since the load is continuously changing, different magnitudes of reactive power jumps were recorded as presented in Fig. 6.7. At the same time, the installed DG starts injecting reactive current a few time steps after the voltage disturbance, as observed in cluster 2 of Fig. 6.7, where the voltage change starts decreasing as the reactive power starts increasing. In this context, it is normal to obtain spikes when subtracting the average load response from the i-th load response, as they correspond to different load amount and composition. To remove those spikes, a simple moving average filter is applied. The resulted reactive power responses for generators are shown in Fig. 6.11.

Then, the filtered reactive power responses are used together with the voltage signals $\Delta V_i = V_i(t) - V_{0,i}$ to estimate a 1st or 2nd order transfer function for each of the reactive power control strategies that were identified within the various ADNs. A simplified schematic diagram of this procedure is shown in Fig. 6.12, where each measurement corresponds to the pair ($\Delta V_i(t)$, $\Delta\hat{Q}_{i,\text{gen}}(t)$) in absolute values. The order of the transfer function with the best fit is selected to model the reactive power response of the generators. Finally, the parameters of each transfer function are estimated using the CONtinuous-Time System IDentification (CONTSID) technique [131].

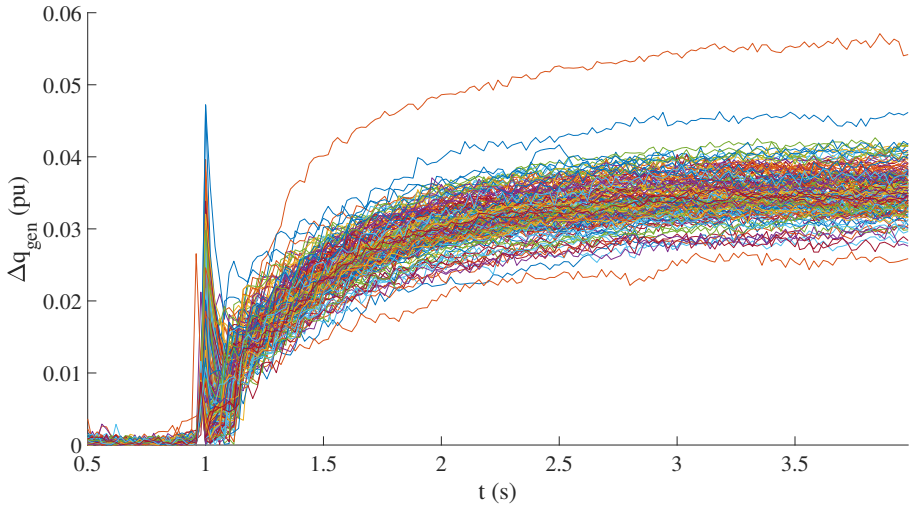


Figure 6.10: Sub.A.1: Estimated reactive power responses for generators in pu.

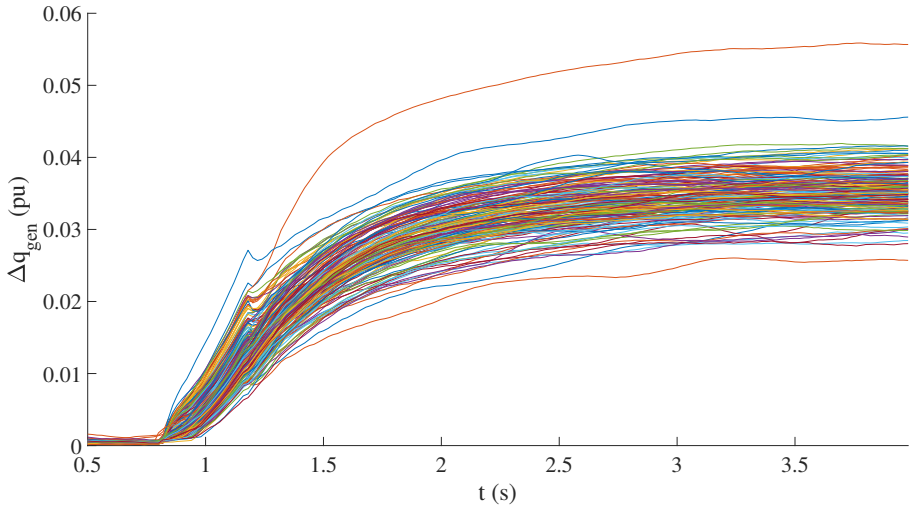


Figure 6.11: Sub.A.1: Estimated reactive power responses for generators in pu after filtering.

System identification results

In this section, the estimated dynamic models of each DG unit are compared. To do so, a common input signal is applied to every transfer function and then, the

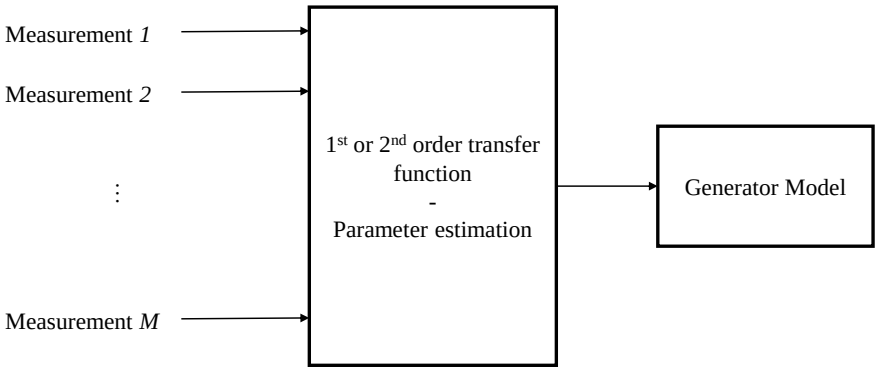


Figure 6.12: System identification for reactive power.

generated dynamic responses are plotted in a common graph. The input signal is a voltage step change of $0.4\,\mathrm{kV}$ as shown in Fig. 6.13. In this context, the generated reactive power responses are illustrated in Fig. 6.14.

It is obvious from Fig. 6.14 that no universal or classical DG unit response can be derived from the measurements. In fact, the reactive power responses are based on the control strategy of each individual DG unit. Therefore, if possible, each DG unit should be modeled explicitly.

For the sake of completeness, Table 6.3 contains the models corresponding to Fig. 6.14. Note that the transfer functions are given in $\mathrm{Mvar/kV}$ and not in per unit. The reason is that, as mentioned above, the provision of reactive power control by DG does not scale proportional to the generation of active power.

In order to use these DG models in the context of a power system model, it is necessary to scale their input and output accordingly. Let $V_{\mathrm{nom,MDN}}$ and $S_{\mathrm{nom,MDN}}$ denote the nominal voltage and power values of the distribution networks whose measurements have been used to obtain the corresponding model. Similarly, let $V_{\mathrm{nom,PSDN}}$ and $S_{\mathrm{nom,PSDN}}$ denote the nominal voltage and power values of the distribution networks that are to be modeled in the power system model.

To obtain the new dynamic responses, the voltage at the PCC of the power

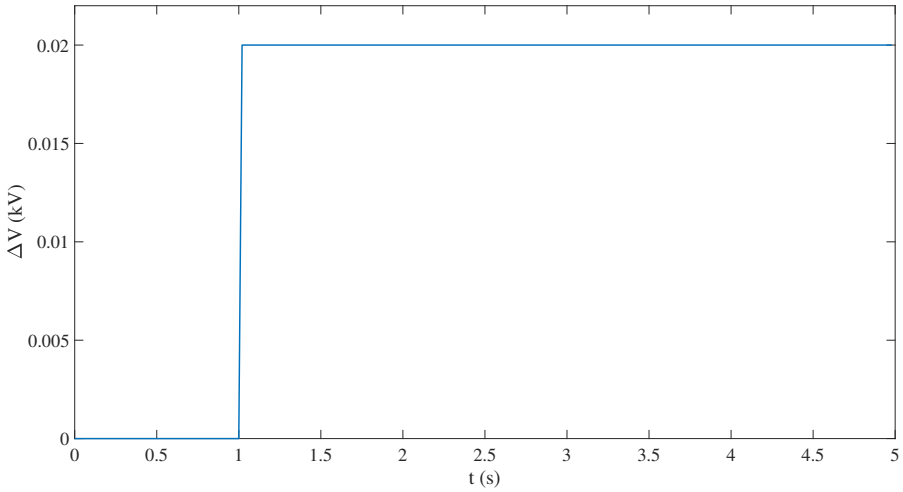


Figure 6.13: Voltage step change of $0.4\,\mathrm{kV}$.

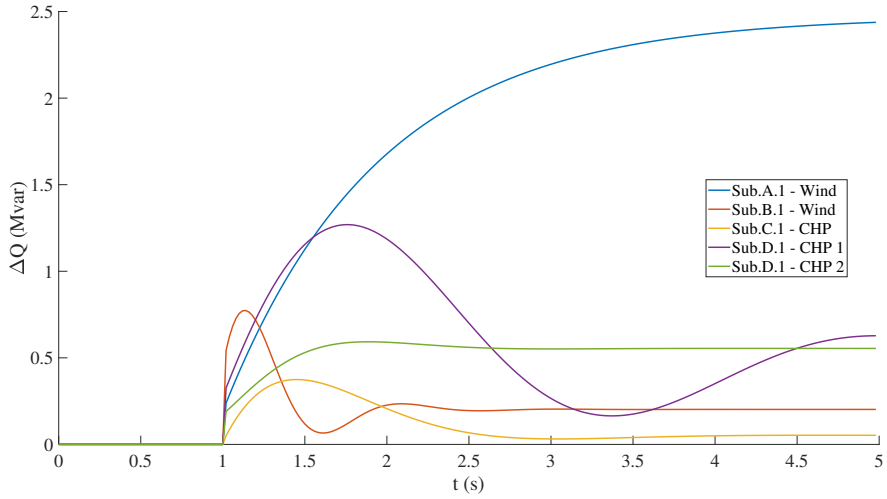


Figure 6.14: Reactive power responses of each generator after a voltage step change of $0.4\,\mathrm{kV}$.

system model should be multiplied by $V_{\mathrm{nom,MDN}}/V_{\mathrm{nom,PDN}}$ before being applied as an input to the DG model. As for the output, it needs to be multiplied by $S_{\mathrm{DG,nom,PSDN}}/S_{\mathrm{DG,nom,MDN}}$. Note that the values for $V_{\mathrm{nom,MDN}}$ and $S_{\mathrm{DG,nom,MDN}}$

are listed in Table 6.3 for each ADN under study.

Table 6.3: Transfer function models $G_{QV,\mathrm{DG}}(s)$ for the reactive power response of dispersed generation.

ADN ID	Nominal Power [MVA]	Nominal Voltage [kV]	Dominating DG	Transfer function $\left[\frac{\mathrm{Mvar}}{\mathrm{kV}}\right]$
Sub.A.1	40	110/20	Wind	$6.18\frac{(1+0.092s)}{1+0.95s}$
Sub.B.1	25	110/20	Wind	$0.51\frac{(1+0.08s)(1+0.64s)}{1+0.83(0.14s)+(0.14s)^2}$
Sub.C.1	40	110/10	CHP	$0.13\frac{(1+0.03s)(1+5.50s)}{1+1.33(0.38s)+(0.38s)^2}$
Sub.D.1	40	110/10	CHP 1	$1.23\frac{(1+0.13s)(1+1.25s)}{1+0.53(0.49s)+(0.49s)^2}$
Sub.D.1	40	110/10	CHP 2	$1.39\frac{(1+1.49(0.16s)+(0.16s)^2)}{1+1.27(0.28s)+(0.28s)^2}$

6.3 Summary

In Chapter 3, it was observed that the active power is not affected by any of the various DG technologies in the ADNs under study. However, this is not a general conclusion, since different control concepts may be implemented in other DG units installed in other distribution grids. Regarding the reactive power, we have observed that it is highly affected by wind parks and CHP plants whereas PV and Biogas plants do not have a direct impact on its dynamic response. Importantly, this is not a general conclusion about reactive power but rather, it is valid only in the examined ADNs. Different ADNs may contain DG with completely different control schemes.

Based on the aforementioned observations about the influence of different DG technologies on system dynamics, several different pu load models were developed for the active power. Those load models correspond to the ADNs where DG data were fully available and the real load demand could be estimated. Using the derived models, we concluded to a generic dynamic model which can be used in dynamic simulations. This generic model was compared with the average model developed several years ago and only small differences in the transient and steady-state responses were observed.

Regarding the reactive power, it was assumed that the loads operate with a constant power factor so that we can derive pu models. For the ADNs with no voltage control, the same methodology as in active power was followed. As a result, several reactive power models were derived corresponding to the pure load dynamics of each ADN under study. Furthermore, a generic load model was derived in order to be used in dynamic simulations for estimating the load's reactive power response during electromechanical transients. This generic model was compared with the average model developed several years ago. Interestingly, no significant difference between the two models was observed.

As for the ADNs with voltage control, the reactive power responses were divided into two respective sub-responses, one caused by the load and one by the implemented control strategy of DG. To do so, the proposed unsupervised learning

method was initially implemented in order to automatically group the recorded data into subsets with and without reactive power control. As a next step, two separate dynamic models were obtained for each of the given ADNs; one for load and one for generation. The load model is developed using the the first subset whereas the generation model using both subsets. In particular, five generation models were developed in order to model the dynamics of the wind parks and CHP plants installed within the examined ADNs.

6.4 Appendix

In this section, the individual generator responses are presented for Sub.B.1, Sub.C.1, and Sub.D.1. Note that those curves correspond to the filtered reactive power responses as derived by the methodology presented in this chapter.

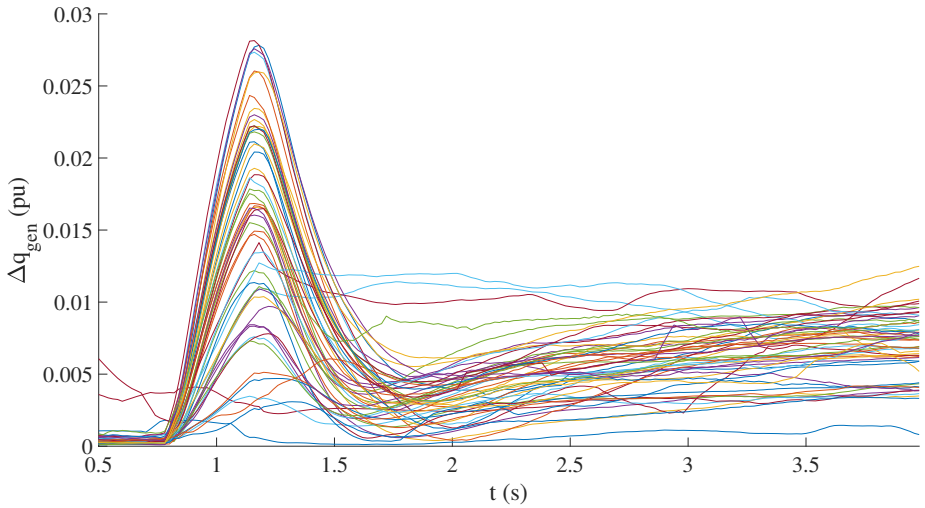


Figure 6.15: Sub.B.1: Reactive power responses for generators in pu after filtering.

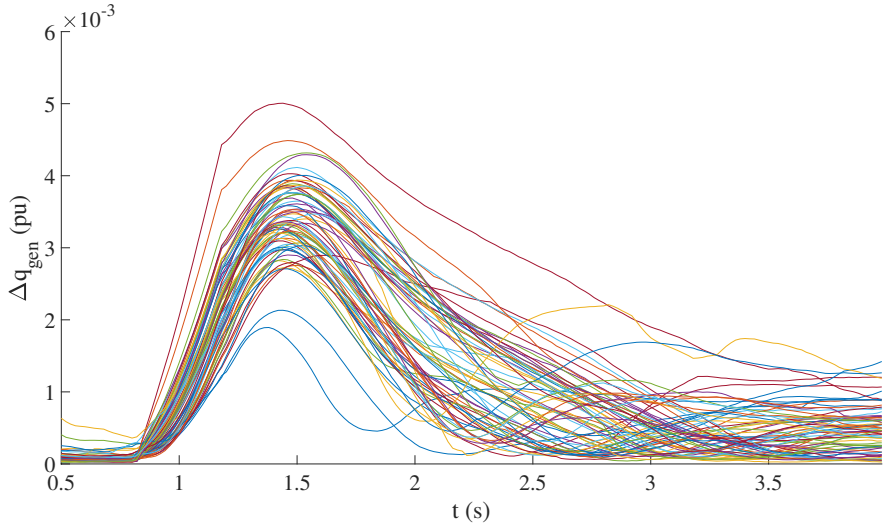


Figure 6.16: Sub.C.1: Reactive power responses for generators in pu after filtering.

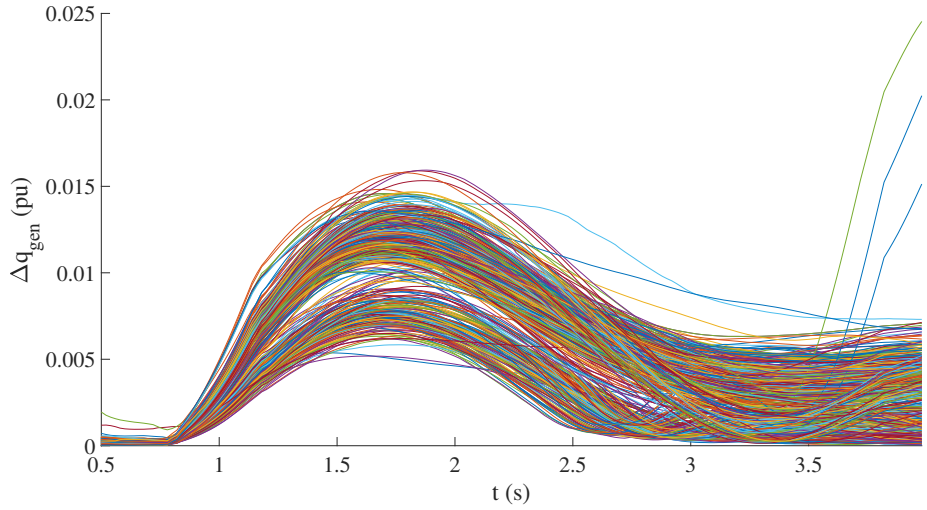


Figure 6.17: Sub.D.1: Reactive power responses for generators in pu after filtering 1.

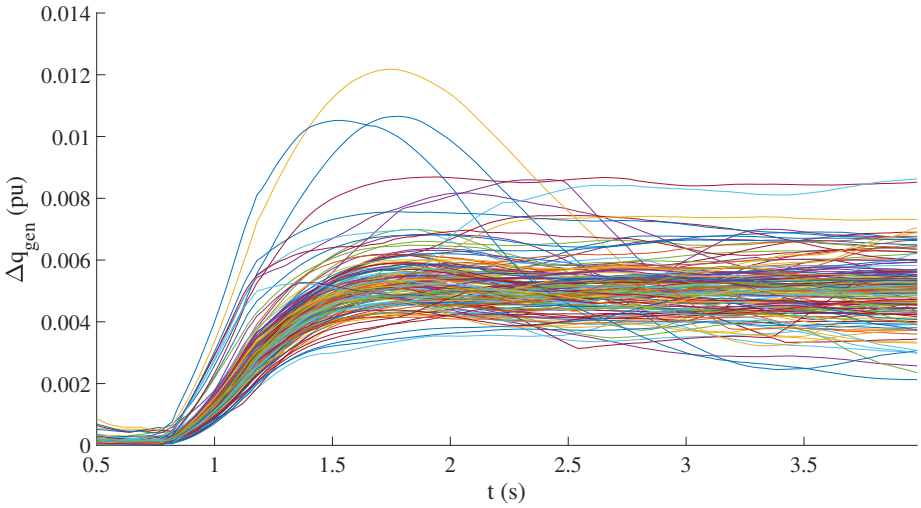


Figure 6.18: Sub.D.1: Reactive power responses for generators in pu after filtering 2.

Chapter 7

Nonlinear Dynamic ADN Model

This chapter starts with a thorough description of the well-established exponential recovery model (ERM) and continues with a discussion about its limitations. Then, a recent modification of ERM is presented, which aims at addressing the limitation of ERM in modeling reverse power flows [4]. Those reverse power flows may occur in modern ADNs with significant DG. Next, the main issue of ERM and its modified version of [4] is identified and discussed. To tackle this issue, a new nonlinear dynamic ADN model structure is proposed. Finally, the efficacy of the proposed model to alleviate the identified issue is assessed using the real data acquired within the scope of this work. Part of the work presented in this chapter can be also found published in [12, 41].

7.1 Exponential recovery model (ERM)

As emphasized in Chapter 1, dynamic equivalent load models are vital elements for simulating and analyzing the dynamics of a power system and thus, they are principally used by TSOs for their system's stability studies. In this context, the ERM has been widely deployed in the literature for simulating distribution networks comprising only load components [39, 67, 43, 132]. However, distribution networks are evolving from passive to active systems containing controllable loads and DG.

Under those conditions, the applicability of the ERM needs to be thoroughly tested and assessed.

As a general rule, the active and reactive power responses of a distribution network after a voltage step are qualitatively similar [67] and they follow a trajectory comparable to the one illustrated in Fig. 7.1. Note that the response in Fig. 7.1 is indicative and the responses in active and reactive power are not assumed to be identical. In principle, those dynamic responses comprise three phases: 1) the transient part where the power changes immediately from Y_0 to Y_{tr} after the voltage step from V_0 to V_{tr}, 2) the recovery part, where the power progressively recovers from Y_{tr} to Y_{ss} and 3) the steady-state part, where the power has reached a new steady-state Y_{ss} for a new steady-state voltage level V_{ss}. Following [67], the mathematical representation of that behavior can be written as

$$Y_{\mathrm{s}}(t) = Y_0(V(t)/V_0)^{N_{\mathrm{s}}}, \tag{7.1}$$

$$Y_{\mathrm{t}}(t) = Y_0(V(t)/V_0)^{N_{\mathrm{t}}}, \tag{7.2}$$

$$Y_{\mathrm{r}}(t) = \mathcal{L}^{-1}\left[(Y_{\mathrm{s}}(s) - Y_{\mathrm{t}}(s))\, G(s)\right], \tag{7.3}$$

$$G(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + ... + b_0}{s^n + a_{n-1}s^{n-1} + ... + 1}, \tag{7.4}$$

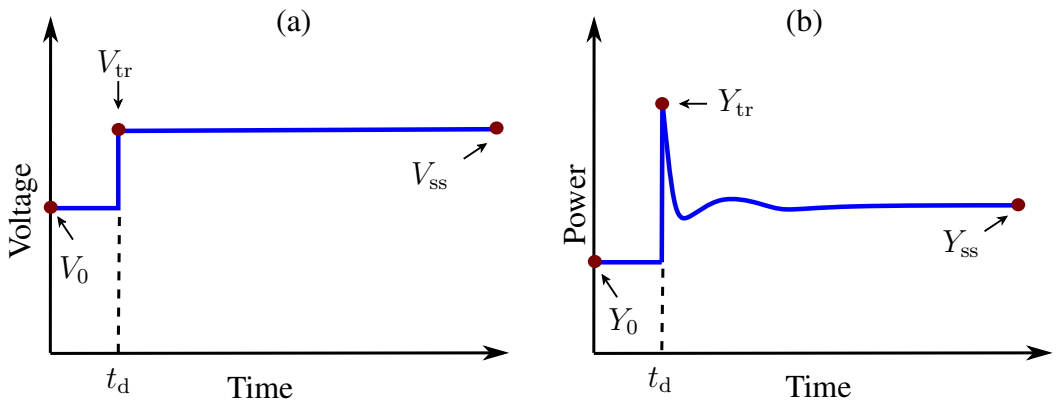


Figure 7.1: Qualitative active or reactive power response of an ADN after a step voltage change: a) Possible voltage step V; b) Indicative (active or reactive) power response Y.

$$\hat{Y}(t) = Y_{\mathrm{r}}(t) + Y_{\mathrm{t}}(t), \tag{7.5}$$

where t is the time, $V(t)$ indicates the bus voltage at each time step, $Y_{\mathrm{s}}(t)$ and $Y_{\mathrm{t}}(t)$ are two nonlinear functions which model the power's steady-state and transient jump, respectively, using the exponents N_{s} and N_{t}. $Y_{\mathrm{s}}(s)$ and $Y_{\mathrm{t}}(s)$ represent the Laplace transform of $Y_{\mathrm{s}}(t)$ and $Y_{\mathrm{t}}(t)$, respectively, $G(s)$ is the variable order transfer function used to approximate the recovery part $Y_{\mathrm{r}}(t)$, and $\hat{Y}(t)$ denotes the model output.

As a first step of the parameter estimation procedure, the exponents N_{s} and N_{t} are calculated based on the acquired measurement using

$$N_{\mathrm{s}} = \log\left(Y_{\mathrm{ss}}/Y_0\right) / \log\left(V_{\mathrm{ss}}/V_0\right), \tag{7.6}$$

$$N_{\mathrm{t}} = \log\left(Y_{\mathrm{tr}}/Y_0\right) / \log\left(V_{\mathrm{tr}}/V_0\right). \tag{7.7}$$

Subsequently, if any of Y_{ss}, Y_{tr}, and Y_0 have different signs, the model exponents cannot be determined, since logarithmic functions are not defined for negative arguments. Hence, the modeling of an ADN, in which the power can change sign after a voltage disturbance, is not feasible with (7.1)–(7.5).

7.2 Modification of exponential recovery model

As underlined in the previous section, the ERM employs two distinct exponents to model the steady-state and transient power response. However, this formulation is not suitable to cover bi-directional power flows that may occur in the context of active distribution networks (ADNs). To tackle this issue, a modification of the ERM has been proposed replacing the two exponential functions with two first-order polynomial functions [4].

In order to tackle the aforementioned issue, [4] proposes the substitution of (7.1) and (7.2) with the first-order polynomial functions

$$Y_{\mathrm{s}}(t) = Y_0\left[\alpha_1\left(V(t)/V_0\right) + \alpha_2\right], \qquad \alpha_1 + \alpha_2 = 1, \tag{7.8}$$

$$Y_{\mathrm{t}}(t) = Y_0\left[\beta_1\left(V(t)/V_0\right) + \beta_2\right], \qquad \beta_1 + \beta_2 = 1, \tag{7.9}$$

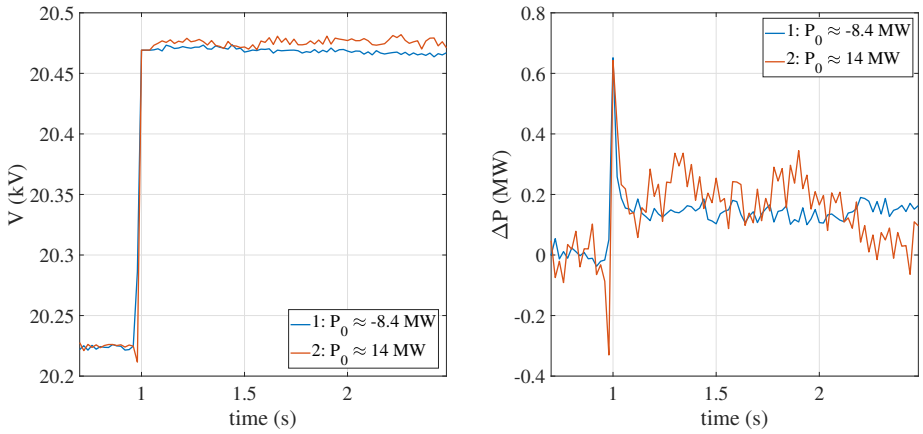
while the rest of the model remains unchanged.

To estimate the model parameters, the initial, steady-state and transient values of power and voltage are deployed. However, those values cover only one operating point of the examined system. In this context, although the first-order polynomial functions may be able to approximate the responses generated by a nonlinear exponential function around an equilibrium point, they may fail to accurately capture the nonlinear characteristics of an ADN at another operating point.

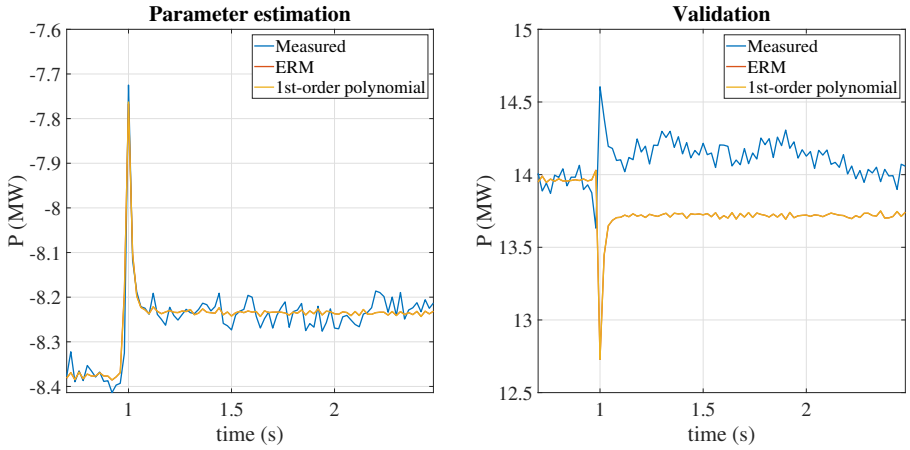
In addition, both ERM model and its modification rely on the measured power at the PCC, which expresses the net consumption and does not reflect the real power demand due to DG. However, both models scale their outputs based on Y_0, as shown in (7.1), (7.2), (7.8), and (7.9). Hence, they may fail to accurately simulate new test scenarios of the same system under different levels of DG. Moreover, in case that the model parameters have been estimated using an operating point with positive Y_0 and a new test scenario with negative Y_0 needs to be simulated (or vice versa), both models will generate a dynamic response towards the wrong direction for the same voltage disturbance.

To highlight this issue, two identical voltage disturbances and their respective change in active power are presented in Fig. 7.2a, as recorded in one of the examined substations. The change in active power ΔP denotes the actual difference between the recorded P signal and its corresponding initial power demand P_0. As depicted in the Fig. 7.2a, apart from the different noise levels, the two dynamic responses are very similar even though the two recorded P_0 differ greatly, i.e., $-8.4\,\mathrm{MW}$ and $14\,\mathrm{MW}$. In this context, we deploy the voltage and active power signals of the first measurement ($P_0 \approx -8.4\,\mathrm{MW}$) to build an ERM and a modified ERM according to [4]. Then, we validate the yielded models using the respective signals of the second measurement ($P_0 \approx 14\,\mathrm{MW}$). As illustrated in Fig 7.2b, both models can accurately simulate the response of the system for the first case whereas they fail to generalize for the second case.

A possible solution to tackle this issue would be to adjust Y_0 based on the active power of DG in order to estimate the real power consumption $Y_{0,\mathrm{real}}$ for each event. However, the estimation of $Y_{0,\mathrm{real}}$ requires high resolution data of the gener-



(a) Two recorded voltage disturbances with similar active power responses.



(b) Output of ERM and its modification for training and validation data. Both responses are identical, and thus the red curve is covered by the yellow one.

Figure 7.2: Limited generalization capability of ERM and modified ERM (1st-order polynomial).

ated active power by the DG units. Privacy policies may not allow DSOs to share those data with TSOs [11]. Furthermore, DSOs may only have access to metering DG data, as communicated by two big DSOs. Due to the high volatility of renewable DG, metering data with a resolution of 15 min are insufficient for a reliable estimation of $Y_{0,\text{real}}$.

7.3 Proposed nonlinear dynamic model

Both ERM and its modification in [4] rely on the initial pre-disturbance power Y_0 in order to calculate the steady-state and transient terms. However, this value does not describe the real power demand in modern ADNs with significant DG. DG does not only offset the recorded active power consumption but the DG units may also inject/absorb reactive power in order to provide voltage support. Therefore, under high DG penetration the recorded Y_0 at PCC may yield misleading dynamic responses. To this end, a modification of the ERM is proposed, deploying the actual differences of the recorded signals $V(t)$ and $Y(t)$ as

$$\Delta V(t) = V(t) - V_0 \quad \text{and} \quad \Delta Y(t) = Y(t) - Y_0. \tag{7.10}$$

It is reminded that the variable Y can describe either the active or reactive power while (V_0, Y_0) can be considered any operating point within a few time steps before the disturbance.

7.3.1 Proposed model structure

The proposed dynamic model relies on the simple principle that during the steady-state and transient phases, a voltage rise ($\Delta V(t) \geq 0$) generates a positive power change ($\Delta Y(t) \geq 0$) whereas a voltage drop ($\Delta V(t) \leq 0$) results in a negative power change ($\Delta Y(t) \leq 0$) [43]. Therefore, the sign of the voltage change determines the direction of the steady-state and transient power response. Furthermore, the initial power Y_0 does not reflect the real load consumption due to DG, and thus its direct deployment into the system dynamics is avoided. Based on those two intrinsic properties of an ADN, the following dynamic model is proposed, which captures the nonlinear characteristics of a distribution network after a voltage disturbance:

$$h_1\left(V(t)\right) = \Delta Y_{\mathrm{s}}(t) = \operatorname{sgn}\left(\Delta V(t)\right) \cdot a_{\mathrm{s}} \cdot \left|\Delta V(t)\right|^{N_{\mathrm{s}}}, \tag{7.11}$$

$$h_2\left(V(t)\right) = \Delta Y_{\mathrm{t}}(t) = \operatorname{sgn}\left(\Delta V(t)\right) \cdot a_{\mathrm{t}} \cdot \left|\Delta V(t)\right|^{N_{\mathrm{t}}}, \tag{7.12}$$

$$Y_{\mathrm{r}}(t) = \mathcal{L}^{-1}\left[\left(\Delta Y_{\mathrm{s}}(s) - \Delta Y_{\mathrm{t}}(s)\right) \cdot G(s)\right], \tag{7.13}$$

$$\Delta\hat{Y}(t) = Y_{\mathrm{r}}(t) + \Delta Y_{\mathrm{t}}(t), \tag{7.14}$$

$$\hat{Y}(t) = \Delta\hat{Y}(t) + Y_0, \tag{7.15}$$

where $\Delta Y_{\mathrm{s}}(t)$ and $\Delta Y_{\mathrm{t}}(t)$ describe the steady-state and transient power response relative to Y_0 while $a_{\mathrm{s}}, N_{\mathrm{s}}$ and $a_{\mathrm{t}}, N_{\mathrm{t}}$ denote a scaling and an exponential coefficient for the steady-state and transient response, respectively. $\Delta Y_{\mathrm{s}}(s)$ and $\Delta Y_{\mathrm{t}}(s)$ indicate the Laplace transform of $\Delta Y_{\mathrm{s}}(t)$ and $\Delta Y_{\mathrm{t}}(t)$, respectively; their difference is the input of the variable order transfer function $G(s)$ that models the load recovery part $Y_{\mathrm{r}}(t)$. Finally, $\Delta\hat{Y}(t)$ expresses the complete dynamic response of the system relative to Y_0, while the final power output $\hat{Y}(t)$ is reconstructed by simply adding the relative dynamic response $\Delta\hat{Y}(t)$ and the initial power value Y_0. The block representation of the system is depicted in Fig. 7.3.

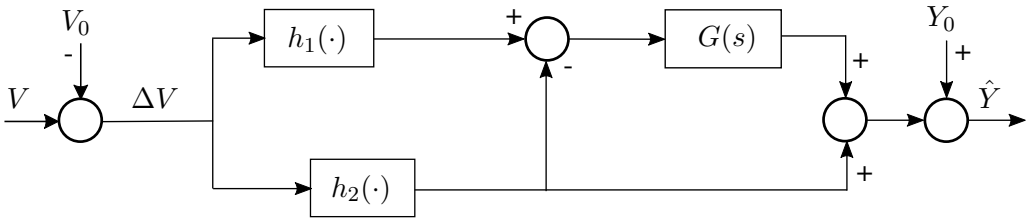


Figure 7.3: Block representation of the proposed nonlinear dynamic model.

7.3.2 Parameter estimation

As pointed out in Chapter 5, a different dynamic model is to be developed for each of the derived clusters. Therefore, the aim of this work is to build generic dynamic models for each individual cluster k. Given that each cluster contains measurements with similar dynamic behavior, a generic set of model parameters $\boldsymbol{\theta}_k = [a_{\mathrm{s},k}, N_{\mathrm{s},k}, a_{\mathrm{t},k}, N_{\mathrm{t},k}, \boldsymbol{p}_k]$ is estimated by means of the corresponding cluster entries. The vector $\boldsymbol{p}_k$ denotes the coefficients of $G_k(s)$.

The representative steady-state and transient parameters of (7.11) and (7.12) can be estimated using nonlinear least squares (NLS) over the M_k cluster entries as:

$$\min_{a_{\mathrm{s},k}, N_{\mathrm{s},k}} \sum_{i=1}^{M_k} \left(\Delta Y_{\mathrm{ss},i} - a_{\mathrm{s},k} \cdot \Delta V_{\mathrm{ss},i}^{N_{\mathrm{s},k}}\right)^2, \tag{7.16}$$

$$\min_{a_{\mathrm{t},k},N_{t,k}} \sum_{i=1}^{M_k} \left(\Delta Y_{\mathrm{tr},i} - a_{\mathrm{t},k} \cdot \Delta V_{\mathrm{tr},i}^{N_{\mathrm{t},k}}\right)^2, \tag{7.17}$$

where

$$\Delta Y_{\mathrm{ss},i} = \left|Y_{\mathrm{ss},i} - Y_{0,i}\right|, \tag{7.18}$$

$$\Delta Y_{\mathrm{tr},i} = \left|Y_{\mathrm{tr},i} - Y_{0,i}\right|, \tag{7.19}$$

$$\Delta V_{\mathrm{ss},i} = \left|V_{\mathrm{ss},i} - V_{0,i}\right|, \tag{7.20}$$

$$\Delta V_{\mathrm{tr},i} = \left|V_{\mathrm{tr},i} - V_{0,i}\right|. \tag{7.21}$$

For the parameter estimation, a genetic algorithm is deployed in order to find the global minimum of (7.16) and (7.17) [133].

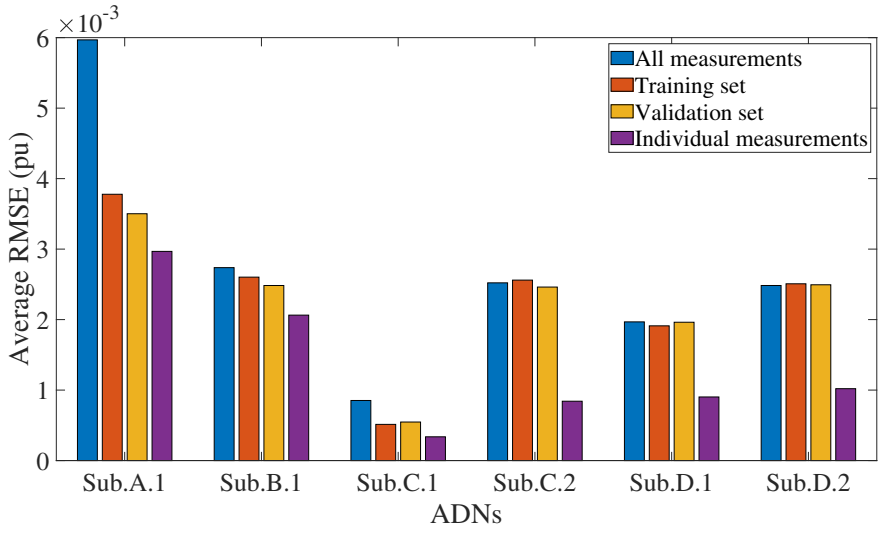
Regarding $G_k(s)$, a first, a second, and a third order transfer function are estimated using the continuous-time system identification (CONTSID) technique [134]. The one that yields the best fit in terms of normalized mean square error between the simulated responses and the real output values is selected. The order of $G_k(s)$ was intentionally chosen low in order to capture the loads' general characteristics while avoiding overfitting. Besides that, the measurements acquired in six real substations did not reveal higher order dynamics.

7.4 Results

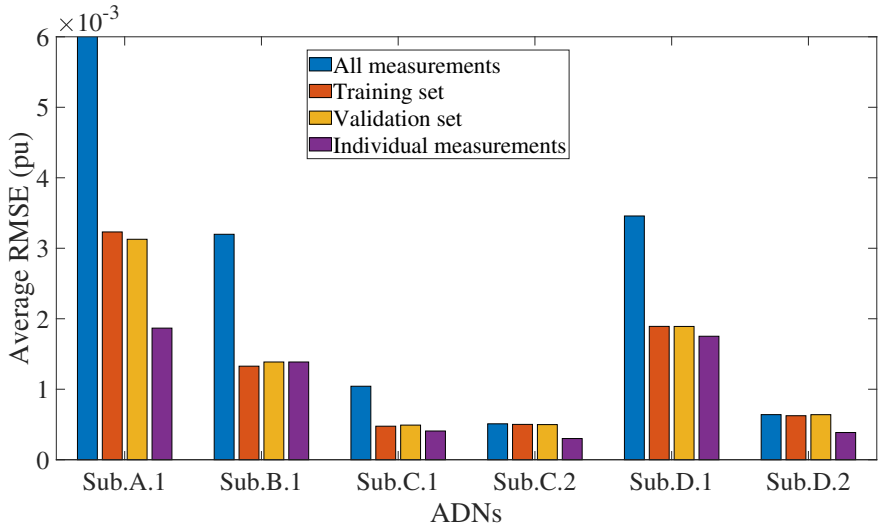
7.4.1 Generalization capability of the clusters

Once the clusters have been formed, a dynamic model is derived for each cluster. To do so, 70% of each cluster's entries are randomly selected as a training set to estimate the model parameters and the rest 30% is used as a validation set. To assess the accuracy of the model, the RMSE between the model output and the real measurement values is estimated for both the training and validation set. In addition, in order to evaluate the generalization capability of the various cluster models, their RMSEs are compared with the ones yielded in the ideal case in which each measurement is fitted into a separate model. On the other hand, in order to

quantify the clustering gain in the modeling accuracy, only one model is developed for each ADN using all its measurements and its RMSE is compared with the ones yielded by the respective clusters.



(a) Bar graphs of RMSE for active power.



(b) Bar graphs of RMSE for reactive power.

Figure 7.4: Bar graphs of the average RMSE between the respective model output and the real measurements for all ADNs.

In Fig. 7.4, the average RMSEs of the three different modeling approaches are

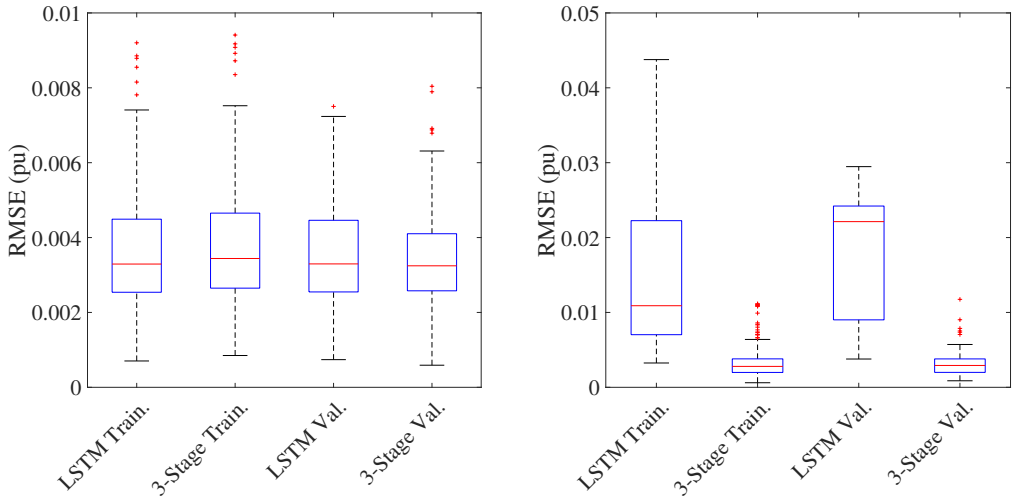
presented, i.e., one model per ADN fitted with all the recorded data, one model per cluster and one model per individual measurement. Those plots correspond to six representative ADNs, as identical plots were generated for the rest. In case of active power, splitting the measurements into clusters does not show any advantage (in the examined ADNs), since identical RMSEs can be yielded by only one dynamic model. Importantly, one model seems to be able to capture the general ADN characteristics as its accuracy do not differ significantly from modeling each measurement separately. Contrary to active power, the modeling accuracy of reactive power is increased remarkably using the clustering approach, having almost the same results with the individual modeling. In Sub.A.1, Sub.B.1, Sub.C.1, and Sub.D.1, where there is DG installed and dynamic voltage support is provided through reactive current injections, the need for splitting the measurements into clusters is imperative, as it is confirmed by the modeling results. Sub.C.2 and Sub.D.2 are described by only one cluster, and thus no such increase in accuracy is observed.

Here, it should be noted that the individual models overfit the data and may also factor in consumption trends as well as the existing noise. Hence, their use in dynamic simulations should be avoided even though their accuracy is high. Also, it would not be clear which model to select for dynamic simulations, as there would be so many models as the total number of measurements. Instead, the modeling based on clusters can capture well the respective ADN dynamics without overfitting the data, as it is verified by comparable results for both training and validation set.

7.4.2 Comparison with a deep learning approach

In recent years, the advances in deep learning have tremendously affected many research areas including the development of equivalent models for (active) distribution networks. As for the latter, the majority of the existing work targets purely load-composed distribution networks as also pointed out in the introduction. Nevertheless, in the recent work of [3], an LSTM RNN architecture is introduced in order to tackle ADN modeling. In this section, we compare this approach with our proposed three-stage methodology as a whole. To do so, the RNN model is formed using

50 LSTM blocks and one fully connected layer while the full recorded datasets are randomly split into training and validation sets. The training is performed using the Adam optimizer [135]. To thoroughly evaluate the performance of each approach, we calculate the RMSE between the final model output and the real recorded values. Six indicative boxplots of the yielded RMSEs for Sub.A.1, Sub.D.1, and Sub.F.2 are illustrated in Fig. 7.5-7.7.

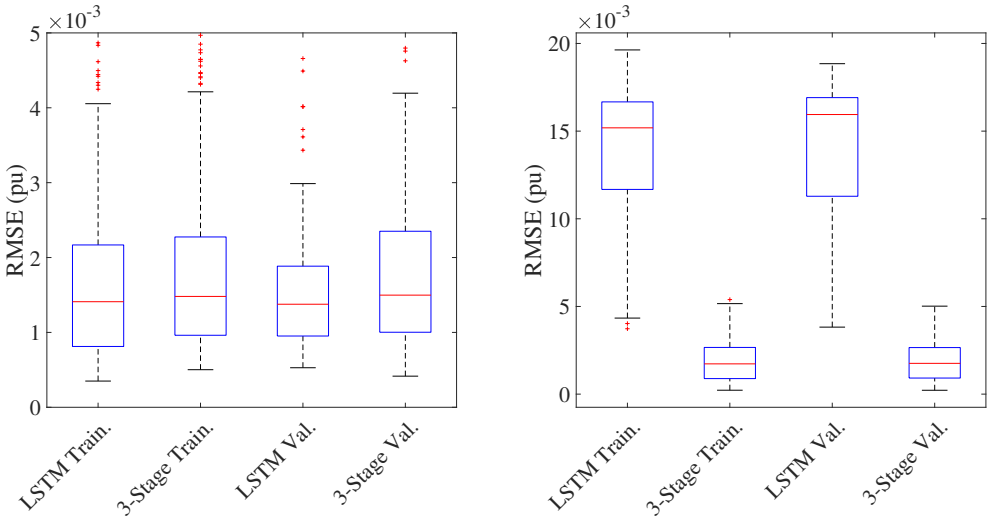


(a) Boxplots for active power.

(b) Boxplots for reactive power.

Figure 7.5: Sub.A.1: Boxplots of the yielded RMSEs using the proposed three-stage methodology and the deep learning one of [3].

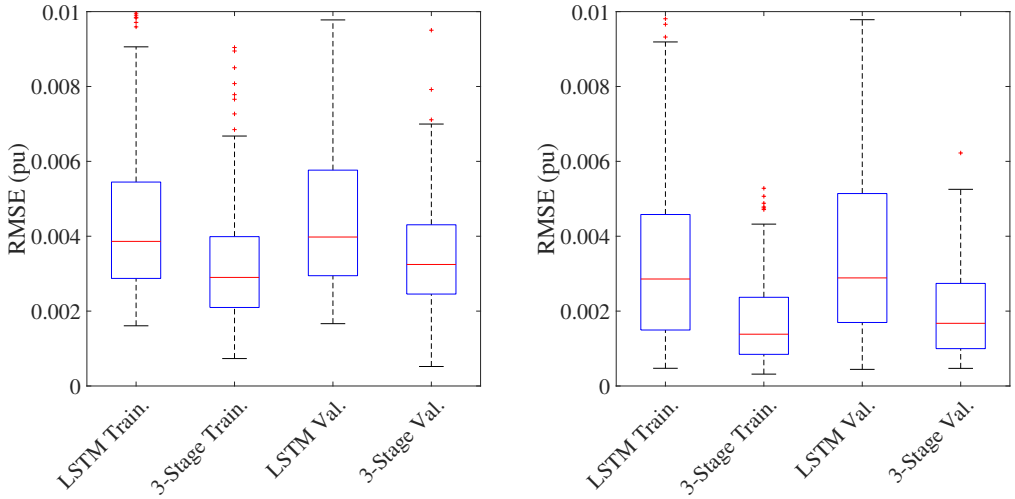
Regarding active power, both approaches, i.e., proposed method and LSTM, generated similar results for almost all the examined ADNs. Only in Sub.F.2 (Fig. 7.7a), the proposed three-stage method outperformed LTSMs. Since all ADNs under study experience similar dynamics with respect to active power, i.e., the classical load recovery behavior, the difference in accuracy in Sub.F.2 originates from the share of the recorded irrelevant measurements. For instance, most of the data acquired in Sub.A.1 (Fig 7.5a) are suitable for parameter estimation, whereas Sub.F.2 is characterized by a larger number of irrelevant events. Those data affect the system identification procedure since LSTMs are trained using a part of the full dataset. As a



(a) Boxplots for active power.

(b) Boxplots for reactive power.

Figure 7.6: Sub.D.1: Boxplots of the yielded RMSEs using the proposed three-stage methodology and the deep learning one of [3].



(a) Boxplots for active power.

(b) Boxplots for reactive power.

Figure 7.7: Sub.F.2: Boxplots of the yielded RMSEs using the proposed three-stage methodology and the deep learning one of [3].

result, the model accuracy and generalization capability deteriorate, underlining the necessity of the first "cleaning" stage. This conclusion is also confirmed by the two indicative validation set examples shown in Fig. 7.8 and 7.9, where the LSTM approach can accurately capture the general active power dynamics in Sub.A.1, while it fails to correctly reflect the real transient response in Sub.F.2.

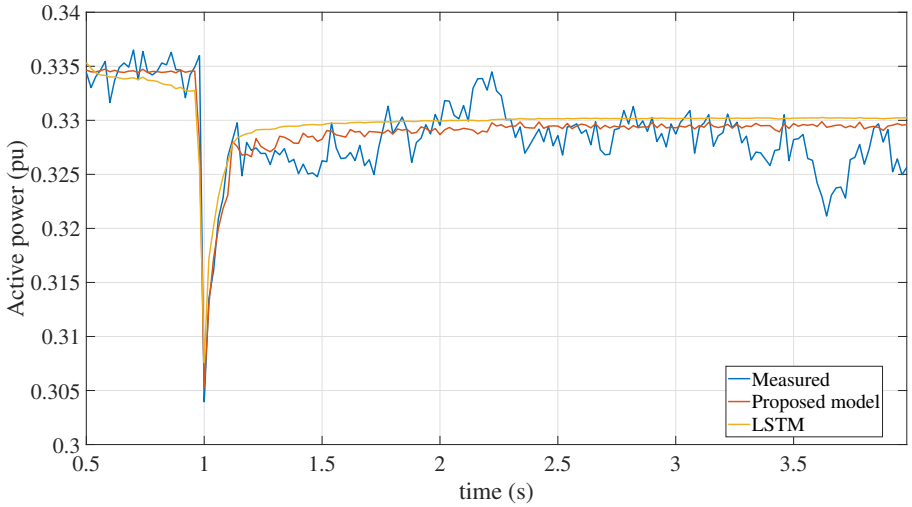


Figure 7.8: Sub.A.1: Individual active power responses.

As for reactive power, it has been proved that the system dynamics can entirely change depending on whether a DG unit operates and provides voltage support through reactive current injections. Under those conditions, it is expected that a model being trained with the full dataset of reactive power signals would not be able to converge to a meaningful set of parameters. This hypothesis is verified by the boxplots of Fig. 7.5b and 7.6b, where a remarkable difference in modeling accuracy is observed between the two approaches. Identical results are obtained for Sub.B.1 and Sub.C.1, which also exhibit voltage support schemes through reactive current injections. In this context, indicative reactive power examples are presented in Fig. 7.10 and 7.11 showing the inability of LSTMs to deal with an heterogeneous dataset. Regarding the ADNs with no voltage support dynamics, the two modeling approaches show negligible difference in accuracy. Only in Sub.F.2, the proposed

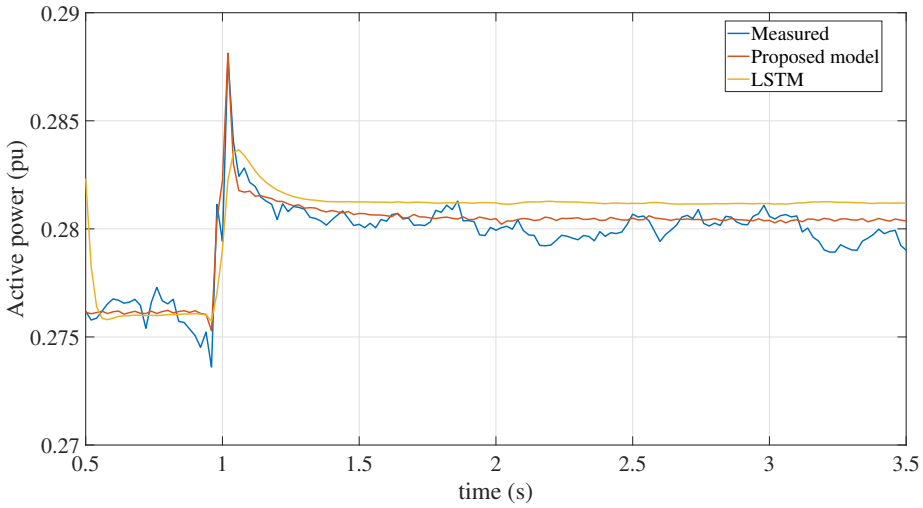


Figure 7.9: Sub.F.2: Individual active power responses.

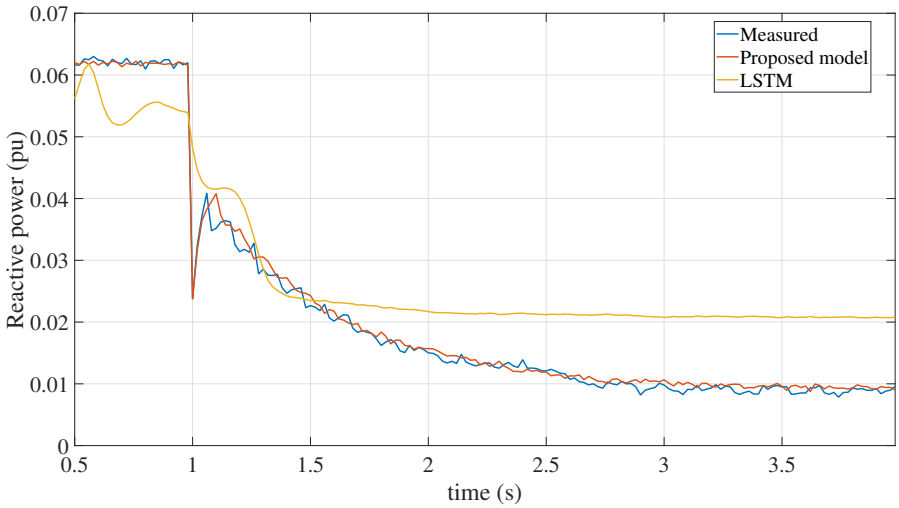


Figure 7.10: Sub.A.1: Individual reactive power responses.

method achieved better results than the LSTM approach (as in the case of active power). This result is attributed to the larger number of irrelevant measurements that were recorded at this particular ADN. Those measurements were included in the training of the LSTM model whereas they were completely discarded in the

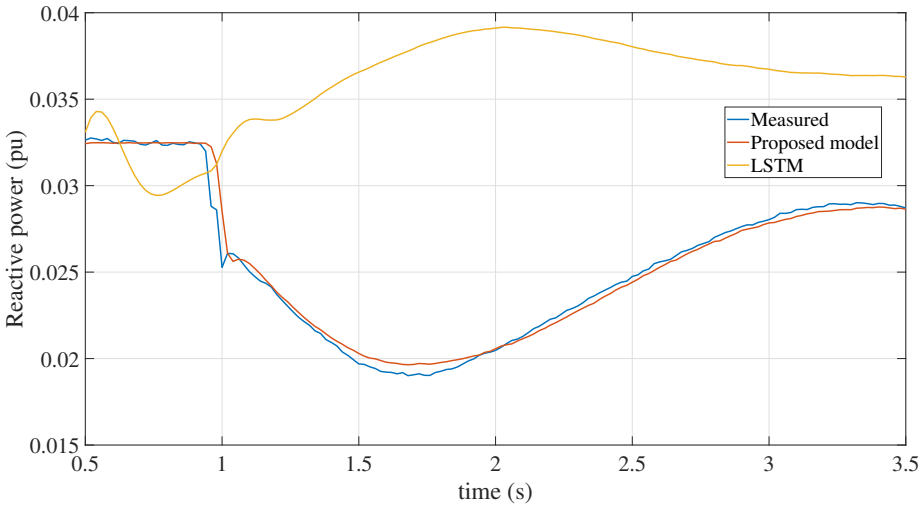


Figure 7.11: Sub.D.1: Individual reactive power responses.

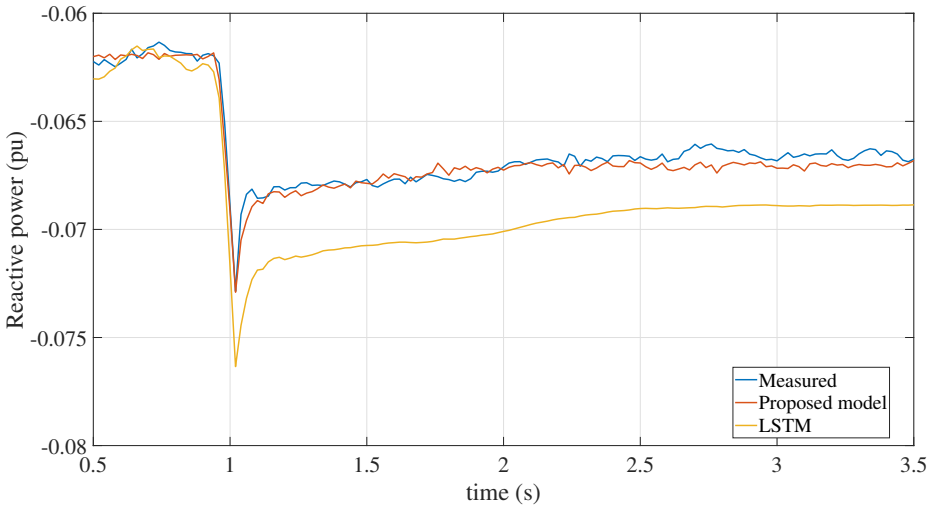


Figure 7.12: Sub.F.2: Individual active power responses.

proposed approach. An indicative example of the generated dynamic responses is presented in Fig. 7.12.

Finally, since the training and the computational time play a critical role in the effectiveness of data-driven techniques, we compare the respective two metrics in

Table 7.1. The training time describes the time needed for the proposed method to estimate the model parameters of all cluster models and for the LSTM to estimate all its parameters using the training set derived by the whole dataset. The computational time denotes the time required for each model to generate its output for an input signal of $100\,\mathrm{Hz}$ sampling frequency and $3.5\,\mathrm{s}$ duration. As expected, the proposed model requires three times less computational time to generate an output response due to significantly less calculations involved. On the contrary, the average training time does not reveal a substantial difference. The training and the generation of the individual responses were performed with MATLAB on a personal computer with an Intel Core i5-8500 CPU, 3.00 GHz processor and 8 GB of RAM.

Table 7.1: Average training and computational time of the proposed model and the LSTM of [3].

	Av. training time	Av. computational time
Proposed model	$52.6\,\mathrm{s}$	$1.6\,\mathrm{ms}$
LSTM	$71.8\,\mathrm{s}$	$5.3\,\mathrm{ms}$

7.4.3 Comparison with other load/ADN models

The scope of this section is to compare the accuracy of the proposed model with three state-of-the-art models; two dynamic models of the same notion and complexity as well as a very recent ADN model based on RNN. In particular, the widely used ERM [67], its most recent modification for modeling ADNs [4] and the LSTM RNN of [3] were implemented using the clustered data of the examined ADNs as yielded by the first two stages of the proposed methodology. Using the same dataset allows for a fair comparison of the model structures.

To compare the accuracy of each model, the RMSE between the respective model output and the field measurement values is calculated for active and reactive power using all the available data. The parameters of ERM and its modification are estimated following the same methodology presented in the parameter estimation

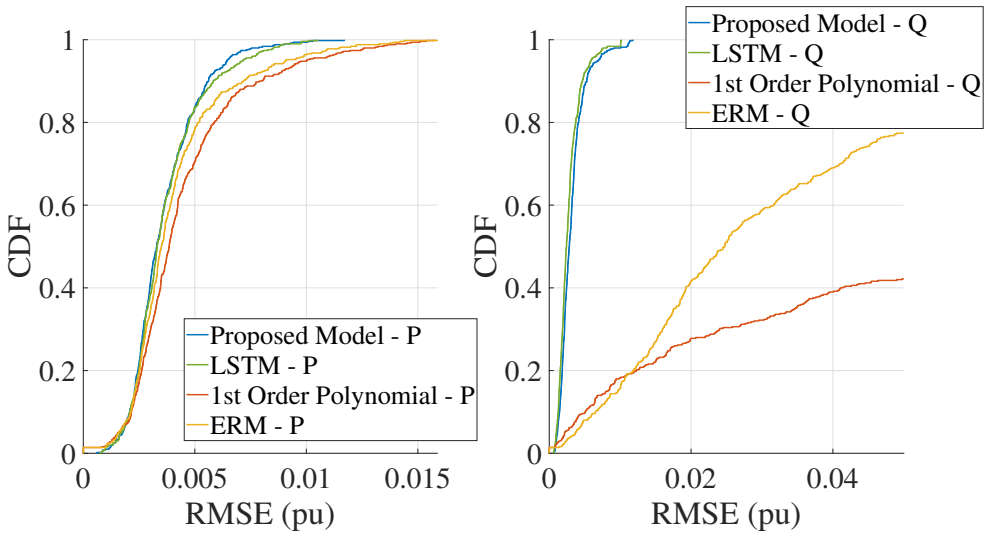


Figure 7.13: Sub.A.1: CDFs of the proposed method, the LSTM of [3], the modified ERM of [4] and the conventional ERM.

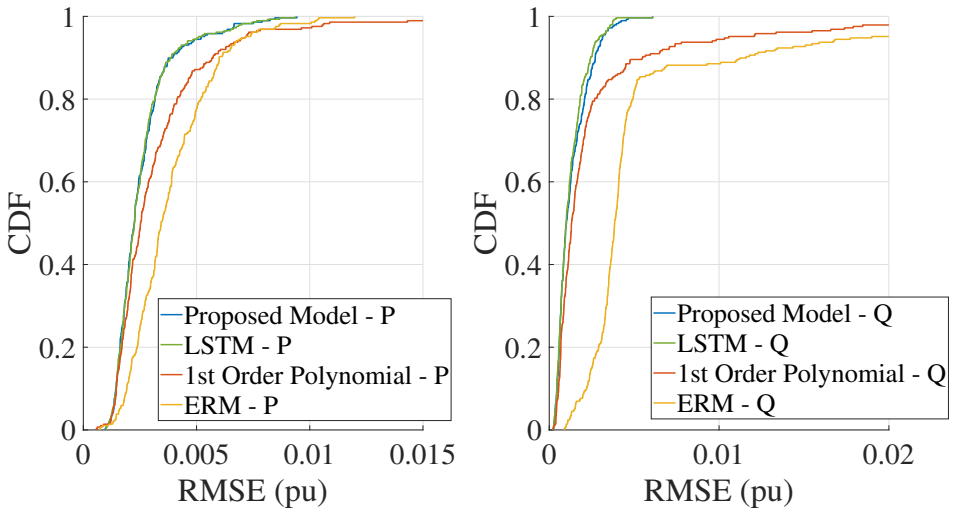


Figure 7.14: Sub.B.1: CDFs of the proposed method, the LSTM of [3], the modified ERM of [4] and the conventional ERM.

section and modifying adequately (7.16) and (7.17), so that they include the corresponding model's steady-state and transient terms. The RNN model was formed

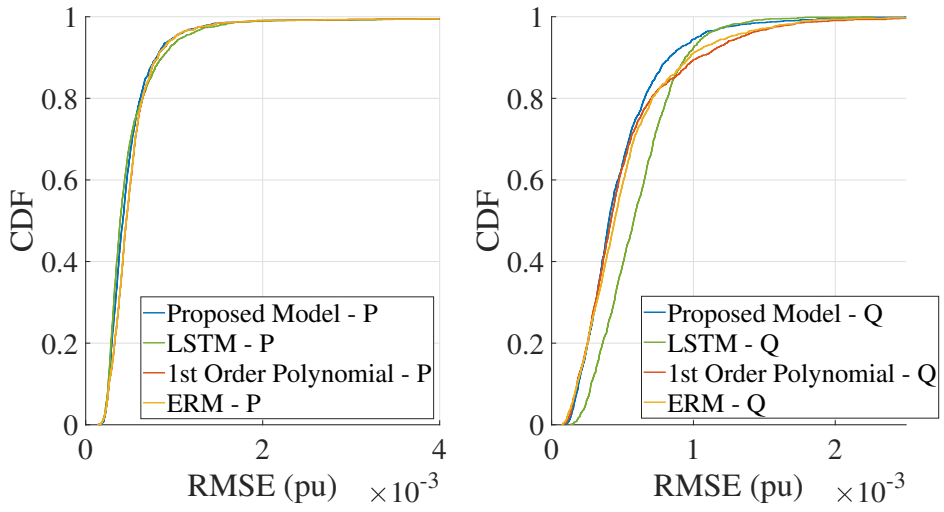


Figure 7.15: Sub.C.1: CDFs of the proposed method, the LSTM of [3], the modified ERM of [4] and the conventional ERM.

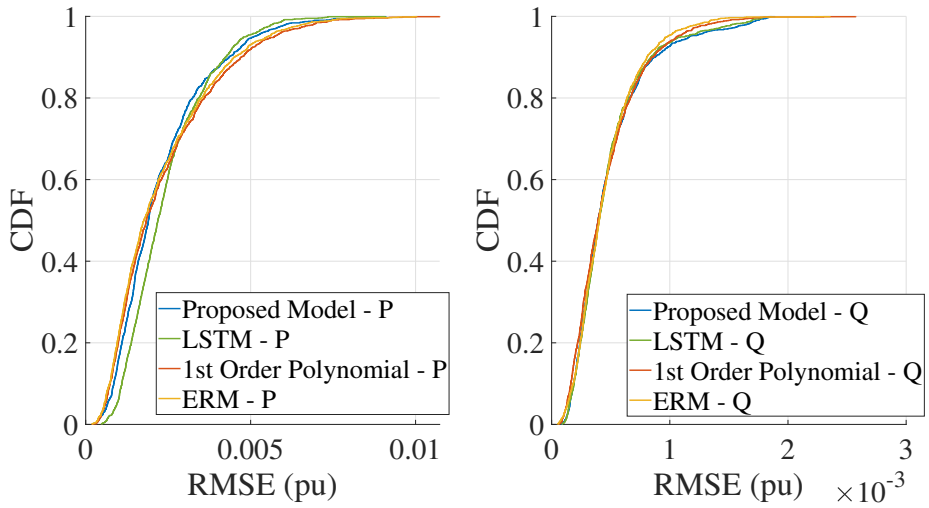


Figure 7.16: Sub.C.2: CDFs of the proposed method, the LSTM of [3], the modified ERM of [4] and the conventional ERM.

using 50 LSTM blocks and one fully connected layer and its training was performed using the Adam optimizer [135]. To thoroughly compare the efficacy of the mod-

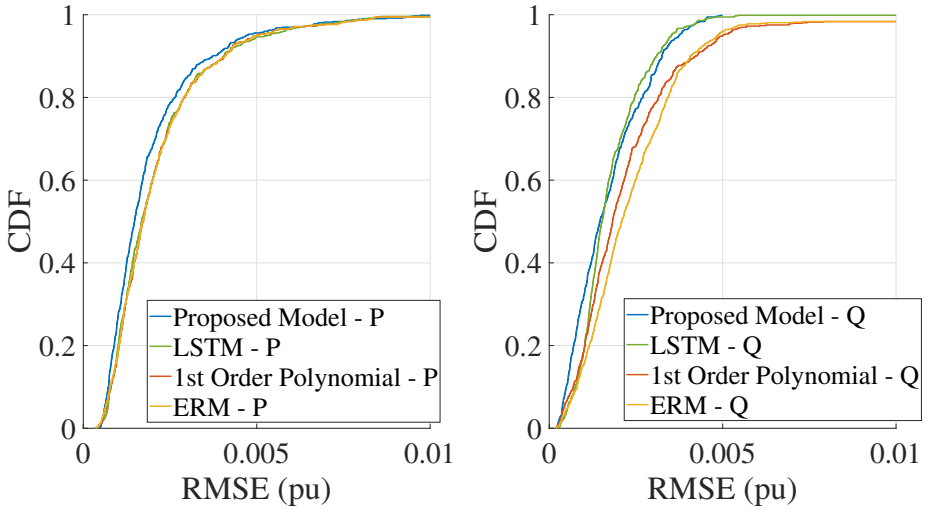


Figure 7.17: Sub.D.1: CDFs of the proposed method, the LSTM of [3], the modified ERM of [4] and the conventional ERM.

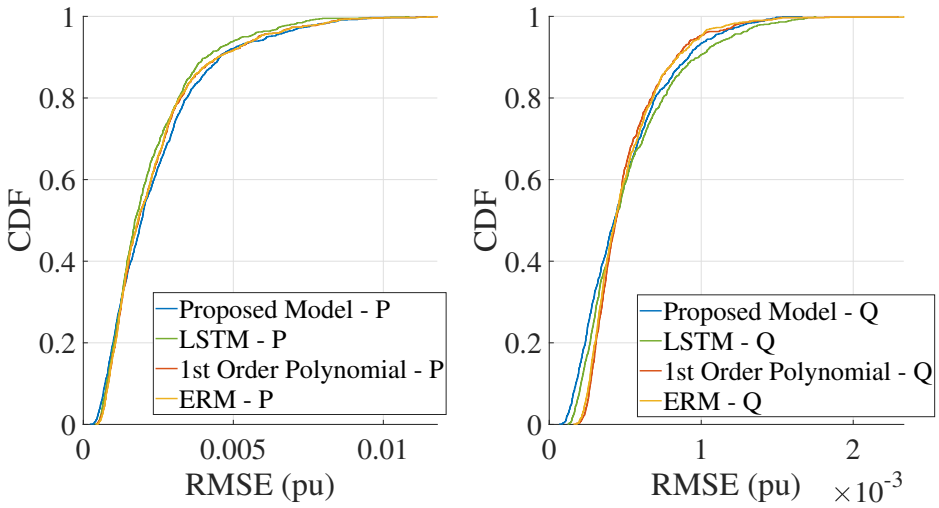


Figure 7.18: Sub.D.2: CDFs of the proposed method, the LSTM of [3], the modified ERM of [4] and the conventional ERM.

els to accurately capture the general ADN characteristics, the cumulative distribution functions (CDFs) of the RMSEs for active and reactive power are exhibited

in Fig. 7.13-7.18 for six representative ADNs. Note that the RMSEs of the various clusters were merged into one group for every ADN and for every load/ADN model, i.e., active and reactive power model.

In all the ADNs, the proposed model yields similar results with the LSTM, with the latter slightly outperforming the former. However, each LSTM comprises around 5000 parameters and requires a longer execution time than the proposed model, which does not exceed the 10 parameters. Regarding the proposed model's comparison with the ERM and the modified ERM, we observed three distinct types of responses. In Sub.A.1, Sub.B.1, Sub.C.1, and Sub.D.1, the proposed model enjoys a significantly higher accuracy, particularly in case of reactive power. This result can be justified by the fact that both ERM and its modification rely on the initial pre-disturbance power Y_0 that does not describe the real power demand in modern ADNs with significant DG. This phenomenon was mostly perceived in Sub.A.1 and Sub.B.1, which are characterized by big wind and PV parks. In contrast, Sub.C.1 and Sub.D.1 contain smaller DG units and thus, their influence on Y_0 is lower. Under those conditions, the proposed model still surpasses the other two models, albeit without a great margin. Finally, all models demonstrate identical RMSEs in case of Sub.C.2 and Sub.D.2, which are composed by pure load elements. Here, the ERM and the modified ERM perform slightly better for around 50% of active power measurements. This small difference in accuracy originates from the fact that the proposed method virtually decouples the dynamic response of the system with its initial conditions in order to make it suitable for ADN modeling. Thus, this missing dependency of the model may lead to small errors when load-composed systems are modeled without however losing its reliability, as validated by the comparison with the other models. In general, the proposed and the LSTM model outperform the two ERMs in ADNs with significant DG, whereas all models show identical results in ADNs with low or no DG.

7.5 Summary

The ERM has been widely deployed in the literature for simulating distribution networks comprising only load components. Howover, distribution networks currently are evolving more and more from passive systems that connect consumers with the rest of the power system to ADNs with significant DG. Under those new conditions, the applicability of ERM was recently questioned. In particular, the ERM employs two distinct exponents to model the steady-state and transient power response. However, this formulation is not suitable to cover bi-directional power flows that may occur in the context of active distribution networks (ADNs).

To tackle this issue, a modification of the ERM has been proposed replacing the two exponential functions with two first-order polynomial functions [4]. Nevertheless, both models rely on the initial power demand at the PCC, which expresses the net consumption and does not reflect the real power demand due to DG. Hence, they may fail to accurately simulate new test scenarios of the same system under different levels of DG. Furthermore, [4] employs affine functions (e.g., first-order polynomials) in order to substitute nonlinear components.

To address those issues, a new formulation of ERM was introduced, which maintains the model's attribute of nonlinearity, captures bi-directional power flows and does not require multiple models to effectively simulate similar dynamics. The proposed equivalent exploits the actual change of the voltage and power signals alongside with two exponential functions. The main novelty of the proposed model lies in its simplicity, generalization capability, and ability to be readily integrated into real-world power system simulation models. Nevertheless, the proposed method cannot capture the dependency between the (unknown) actual load demand and the magnitude of the system's response. Yet the experimental results manifested the applicability of the proposed approach in real-world scenarios.

Moreover, the proposed ERM-inspired nonlinear model structure can be effectively combined with the presented clustering methodology. It was proved that the proposed 3-stage modeling pipeline offers a reasonable balance between accuracy and generalization capability. Under those conditions, the derived ADN models can

be easily integrated into a power system simulation software in order to perform dynamic simulations, e.g., stability studies.

Those results also highlight that, in spite of the theoretically indefinite dynamic models that can describe each combination of load and DG within an ADN, those ADN configurations can actually be mapped into a much smaller data space using cluster analysis. The presented modeling results manifest the capability of the proposed method to capture the general ADN characteristics using only one model per derived cluster without a significant loss in modeling accuracy.

It should underlined that the proposed three-stage methodology can be effortlessly integrated into most of the existing ADN/load modeling approaches. For instance, the performance of a recent deep learning approach based on LSTMs was poor when it was trained using the whole recorded dataset. However, the LSTM model achieved high accuracy results when it was trained using the data of each cluster.

Finally, it was shown that the classical ERM and its modification may lead to less accurate results in cases of ADN with significant DG. This observation originates from the fact that both models directly employ the recorded power value at PCC (Y_0), which, as pointed out, does not reflect the real load demand. For ADNs with low or no DG, those two models can be a reliable choice for stability studies, as they yielded highly accurate results.

Chapter 8

Modeling ADN uncertainty

This chapter starts with an introduction about the need of modeling the stochastic time-varying and weather dependent behavior of loads and DG, in particular in the dynamic load/ADN models. Next, a new way of modeling dynamical systems is presented by leveraging the latest advances in probabilistic machine learning. This modeling approach is then tailored in order to model the dynamics of an ADN. Specifically, the proposed method introduces a probabilistic dynamic model based on Gaussian processes (GPs) that yields a probability distribution for each time step. Due to the inherent flexibility of the model structure, two different alternatives are established. The first one uses the autoregressive input and output terms as features whereas the second one additionally deploys a set of exogenous parameters (apart from the autoregressive terms). Part of the work presented in this chapter can be found in [136, 137].

8.1 Motivation

In general, the measurement-based approach has been proved to work effectively on load and, recently, on ADN modeling. As mentioned earlier in this thesis, it relies on measurement data in order to reconstruct the dynamics of the system without requiring detailed knowledge about the individual system components. To do so,

the parameters of a specific model structure are tuned accordingly so that the model output reflects the recorded signals [17].

However, most of the existing research work employs synthetic measurements artificially generated by a detailed simulation model of the ADN covering its internal components, e.g, [3, 10, 11, 92, 93]. In reality, load and DG may remarkably vary depending on different geographical, temporal, and weather conditions. Therefore, models based on artificially generated data may not be able to fully capture the ADN dynamics in a realistic fashion.

To this end, only a few approaches found in the literature exploit field data acquired in real substations in order to develop the corresponding load or ADN models [24, 25, 26, 27, 28, 38, 39]. Importantly, [25, 26, 39] demonstrated that disparate model parameters may be estimated for different time and weather variations. For instance, different models are developed for 4 different loading conditions using a set of 11 measurements in [25], while 5 different sets of parameters corresponding to different months (from July to November) are estimated in [26]. In the work of [39], measurement data from three summer and three winter days are considered in order to estimate different sets of load model parameters for each season.

Nevertheless, the measurement data deployed in the aforementioned studies are rather limited and do not cover a full year. Therefore, the respective models may not be able to generalize under new test scenarios characterized by different seasonal conditions [40]. Furthermore, those studies were conducted around a decade ago reflecting predominately passive distribution networks with limited or no shares of DG. In this regard, the inspiring work of [40] aims at addressing the issue of limited data by acquiring field measurements for a complete year. However, this approach employs purely static models neglecting the system dynamics while it is restricted to traditional load-composed distribution systems, where no DG dynamics are considered. Furthermore, 384 different models were developed for each substation corresponding to different combinations of half hour intervals within a day (48), the type of the day, i.e., weekend or not (2), and the season (4); a fact that resulted in a large number of nearly identical models. Hence, it becomes clear that a meaningful split of the measurement data in order to build various dynamic models, i.e., devel-

oping models for combinations of hour, day, month, high or low temperature, while avoiding redundant models is not straightforward.

At the same time, most of the inspiring work following the measurement-based approach introduces deterministic dynamic models that do not provide any certainty or confidence about their predictions [3, 11, 13, 25, 30, 33, 37]. However, the power system is characterized by the stochastic nature of loads and DG, where the consumers continuously alter their power demand and the renewable energy sources (RES) adjust their generation based on weather variations. Under those volatile conditions, it is of utmost importance for an ADN model to reflect the system uncertainty induced by load and DG, in particular when previously unseen scenarios are simulated.

Therefore, there are two major open issues in the context of dynamic ADN modeling which are both related to the stochastic properties of the ADN components. To model the uncertainty induced by the stochastic nature of load and DG, a generic probabilistic model based on GPs is proposed. Contrary to the limited approaches fitting random responses into a set of model parameters to account for uncertainty [11], the proposed approach readily yields a probability distribution for each prediction point. Hence, instead of yielding a traditional dynamic response that neglects uncertainty, like the various deterministic models, the proposed model generates confidence intervals alongside with its response. To the best of the author's knowledge, this is the first study introducing a probabilistic dynamic load/ADN model.

In this context, the utilization of the proposed model ranges from predicting the expected dynamic response to reasoning about system uncertainties inherent to measurement data and identifying worst case scenarios using the boundaries of the yielded distributions. In addition, the predictive variance can be also deployed as an indication of whether a prediction should be trusted, further examined or even rejected.

As for the influence of time and weather on load/ADN dynamics, the proposed GP approach is further extended in order to explicitly incorporate a set of exogenous variables, e.g., time of the day, month, temperature, wind speed, etc. To do so, time and a set of publicly available weather parameters are integrated into the model

as random variables. In this context, the model is able to readily reflect time and weather variations while yielding probability distributions as output.

Note that GP was deliberately selected as a model structure for the following reasons:

1. The number of input features can be modified based on the ADN under study.
2. It can accurately model complex nonlinear relationships [138].
3. It is a probabilistic model and thus, it is suitable for encoding the stochastic nature of load and DG.
4. The recorded measurements follow a Gaussian distribution at each time step, as also observed in [11].

8.2 Gaussian processes for modeling dynamical systems

8.2.1 Gaussian process regression

In this section, a short description of the Gaussian process regression is introduced. This step is fundamental for the comprehension of the proposed method. For a more detailed explanation, the interested reader is referred to [138].

As in the classical regression problem, the goal is to learn a function $y = f(\boldsymbol{x})$ given a set of input and output points $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{M}$, where $\boldsymbol{x}_i \in \mathbb{R}^D$ and $y \in \mathbb{R}$. However, instead of assuming a parametric function $y = f_\theta(\boldsymbol{x}; \boldsymbol{\theta})$, a GP regression model assumes that the output comprises a set of finite random variables $f(\boldsymbol{x}_i)$ that follow a joint multivariate Gaussian distribution:

$$f(\boldsymbol{x}_1), ..., f(\boldsymbol{x}_n) \sim \mathcal{N}(0, \mathbf{K}), \tag{8.1}$$

where $\mathbf{K}$ denotes the covariance matrix of this set. In this regard, each element $K_{ij} = C(\boldsymbol{x}_i, \boldsymbol{x}_j)$ represents how much the output at $\boldsymbol{x}_j$ is affected by the output

at $\boldsymbol{x}_i$. Therefore, the covariance function $C(\boldsymbol{x}_i, \boldsymbol{x}_j)$, which is also called kernel, determines the shape of our distribution and can be employed for introducing prior knowledge about the unknown function. Note that the kernel can be formed by any function that results in a positive definite $\mathbf{K}$.

A commonly used covariance function is the radial basis function (RBF), which can be mathematically written as:

$$C_{\text{RBF}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma^2 \text{exp}\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2l^2}\right), \tag{8.2}$$

where σ denotes the standard deviation of f and the parameter l (lengthscale) regulates the decay of correlation between a point and its neighbors.

Assuming that our observations y_i contain Gaussian noise ϵ_i, we can readily express the model output as:

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_n^2), \tag{8.3}$$

where σ_n^2 indicates the variance of the Gaussian noise.

Considering that the parameters of the kernel function and the variance of the Gaussian noise are known, we can estimate the output distribution y^* for a new test input $\boldsymbol{x}^*$. To do so, we write the joint Gaussian distribution as:

$$\begin{pmatrix} \boldsymbol{y} \\ y^* \end{pmatrix} \sim \mathcal{N}(0, \mathbf{K}_{M+1}), \tag{8.4}$$

where

$$\mathbf{K}_{M+1} = \begin{bmatrix} \mathbf{K} & \mathbf{k}(\boldsymbol{x}^*) \\ \mathbf{k}(\boldsymbol{x}^*)^\top & k(\boldsymbol{x}^*) \end{bmatrix}. \tag{8.5}$$

Since Gaussian noise is assumed in our observations, the covariance matrix can be written as $\mathbf{K} = \boldsymbol{\Sigma} + \sigma_n \mathbf{I}$, where $\boldsymbol{\Sigma}$ indicates the covariance matrix of the noise-free system and $\mathbf{I}$ denotes the $M \times M$ identity matrix. Furthermore, $\boldsymbol{y} = [y_1, ..., y_M]^\top$ is an $M \times 1$ vector comprising the observed outputs, $\mathbf{k}(\boldsymbol{x}^*) = [C(\boldsymbol{x}_1, \boldsymbol{x}^*), ..., C(\boldsymbol{x}_M, \boldsymbol{x}^*)]^\top$ is also an $M \times 1$ vector containing the pairwise covariance between the training input points and the new test input $\boldsymbol{x}^*$, and finally $k(\boldsymbol{x}^*)$ is the prior variance of $\boldsymbol{x}^*$.

To estimate the output distribution y^*, the Bayes' rule is applied for the following probability:

$$p(y^*|\boldsymbol{y}, \mathbf{X}, \boldsymbol{x}^*) = \frac{p(\boldsymbol{y}, y^*)}{p(\boldsymbol{y}|\mathbf{X})}, \tag{8.6}$$

where $\mathbf{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_M]^\top$ is the $M \times D$ input matrix. Based on the proof found in [138], the probability $p(y^*|\boldsymbol{y}, \mathbf{X}, \boldsymbol{x}^*)$ follows also a Gaussian distribution with a mean μ and a variance υ that can be described by:

$$\mu(y^*) = \mathbf{k}(\boldsymbol{x}^*)^\top \mathbf{K}^{-1} \boldsymbol{y}, \tag{8.7}$$

$$\upsilon(y^*) = k(\boldsymbol{x}^*) - \mathbf{k}(\boldsymbol{x}^*)^\top \mathbf{K}^{-1} \mathbf{k}(\boldsymbol{x}^*), \tag{8.8}$$

respectively. Due to the fact that we want to model dynamical systems, (8.7) is deployed to yield the dynamic response of the system and (8.8) is interpreted as the confidence level of the predicted dynamic response.

Importantly, the derived vector from the multiplication $\mathbf{k}(\boldsymbol{x}^*)^\top \mathbf{K}^{-1}$ can be considered as a weight vector that is applied to the known output values $\boldsymbol{y}$ in order to predict the unknown y^*. Moreover, in case that the test input $\boldsymbol{x}^*$ positions itself away from the already seen $\{\boldsymbol{x}_i\}_{i=1}^M$, then the multiplication $\mathbf{k}(\boldsymbol{x}^*)^\top \mathbf{K}^{-1} \mathbf{k}(\boldsymbol{x}^*)$ will be small, and thus $\upsilon(y^*)$ will be large. This can be justified by the Bayes' way of thinking, where the less data we see, the greater will be the uncertainty (variance) in the prediction.

Here, it should be noted that the training procedure of a GP lies in estimating the hyperparameters of the kernel function as well as the noise variance based on the observed training set $\mathcal{S}$. If $\boldsymbol{\Theta} = [\sigma, l, \sigma_n]^\top$ is the parameter vector, then a common choice for estimating this vector is the maximization of the log-likelihood [139]:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Theta}) &= \log(p(\boldsymbol{y}|\mathbf{X})) \\ &= -\frac{1}{2}\log(|\mathbf{K}|) - \frac{1}{2}\boldsymbol{y}^\top \mathbf{K}^{-1} \boldsymbol{y} - \frac{M}{2}\log(2\pi). \end{aligned} \tag{8.9}$$

To do so, the derivative of $\mathcal{L}(\boldsymbol{\Theta})$ needs to be computed with respect to each of the parameters as:

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial \Theta_j} = -\frac{1}{2}\mathrm{Tr}\left(\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta_j}\right) + \frac{1}{2}\boldsymbol{y}^\top \mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta_j}\mathbf{K}^{-1}\boldsymbol{y}, \tag{8.10}$$

where Tr indicates the trace. However, each step of the optimization requires the computation of the inverse of **K**, which can become a big computational burden for large training sets.

8.2.2 System identification

As a general rule, GP is largely deployed for solving regression problems. Nevertheless, a significantly smaller amount of research work has extended the applicability of GP to modeling dynamical systems, as shown by, e.g., [140, 141, 142]. In order to model a dynamical system using GP, we need to convert the system identification problem into a regression problem. To do so, a vector of regressors corresponding to the past input and output values is constructed as:

$$\boldsymbol{x}_k = [u(k), ..., u(k-L), y(k-1), ..., y(k-L)], \tag{8.11}$$

where L denotes the order of the model.

To yield a dynamic response, the GP model generates one prediction point $\hat{y}_k$ at a time, which requires the input and output values of the previous time steps, as also indicated by (8.11). In principle, there are two approaches for making multi-step ahead predictions using a GP dynamic model; namely the "naive" and the "exact" approach. The former repeatedly predicts one-step ahead using the mean prediction values of the previous steps. The latter follows the same principle, but instead of the mean prediction values, it leverages the full predicted distributions of the previous steps, i.e., the mean predictions and the corresponding variances [139].

8.3 Dynamic modeling of ADNs using Gaussian processes

ADN dynamics differ notably from hour to hour, day to day and month to month, since load and DG vary not only in magnitude but also in composition [25, 26, 39]. To this end, GPs can be considered as an adequate option to capture those variations, due to their intrinsic probabilistic properties.

8.3.1 Methodology

As a first step, a dataset of M voltage disturbances is acquired at the point of common coupling (PCC). Each recording i is composed by three vectors $\boldsymbol{V}_i$, $\boldsymbol{P}_i$ and $\boldsymbol{Q}_i$ corresponding to the RMS values of voltage, active and reactive power, respectively.

As already pointed out, one of the major differences between conventional distribution networks and modern ADNs is the installed DG. In this context, the recorded active and reactive power at the PCC do not reflect the real power demand of the system, but rather the difference between the real load and the power generated by DG. Therefore, using the raw recorded power values may lead to error-prone models, due to the high volatility of DG, which vary their feed-in depending on the weather conditions and the grid codes for voltage support. To tackle this issue, the system dynamics are virtually decoupled from its initial pre-disturbance conditions by estimating the actual differences of voltage $\Delta V_i(t)$ and power $\Delta Y_i(t)$ as:

$$\Delta V_i(t) = V_i(t) - V_{0,i} \quad \text{and} \quad \Delta Y_i(t) = Y_i(t) - Y_{0,i}. \tag{8.12}$$

The variable Y can indicate either the active or reactive power, while $(V_{0,j}, Y_{0,j})$ can be any operating point a few time steps before the disturbance.

Consequently, we randomly split the acquired measurements into training (70%) and validation (30%) set. Then, we construct the vector of regressors for each time step t and for each acquired measurement i of the training set as:

$$\begin{aligned} \boldsymbol{x}_{i,t} = [&\Delta V_i(t), ..., \Delta V_i(t-L), \\ &\Delta Y_i(t-1), ..., \Delta Y_i(t-L)], \end{aligned} \tag{8.13}$$

which corresponds to the output $y_{i,t} = \Delta Y_i(t)$. Once all the vectors of regressors and the output values have been set, they are merged into a training matrix $\mathbf{X}$ and into an output vector $\boldsymbol{y}$, respectively. Then, $\mathbf{X}$ and $\boldsymbol{y}$ are normalized within the range of $[-1, 1]$. Finally, the RBF kernel (8.2) is selected as a covariance function and its parameters are estimated using (8.9) and (8.10). Importantly, since reliable ADN models require a significant amount of measurements, the direct deployment of all the available data in (8.10) may lead to a computationally intractable problem. To

this end, we deploy only a random subset of 2000 training points at each iteration of the optimization. This can be seen as an analogous technique to the minibatch training in deep learning [143].

Moreover, the order of the system is highly dependent on the dynamics of the examined ADN. For instance, a purely load-composed distribution network may need first or second order dynamics to be accurately modeled. However, an ADN with significant DG, which follows grid codes for voltage support, may require a higher order model. To this end, we suggest a simple methodology in order to determine an adequate model order L. Starting with $L = 1$, the order keeps increasing until the difference in modeling accuracy of the training set between two consecutive model orders does not exceed a user-defined threshold. Then, the training is stopped and L is set to its penultimate value. In this context, the modeling accuracy can be described by the the continuous ranked probability score (CRPS). CRPS is a widely used metric in probabilistic forecasting that considers both the point prediction as well as the predictive uncertainty. It can be mathematically formulated as:

$$\text{CRPS}_i = \int_{-\infty}^{\infty} \left(\hat{F}(y) - \mathbb{1}\left(y - y_i\right) \right)^2 dy, \tag{8.14}$$

where $\hat{F}$ represents the predicted cumulative distribution function (CDF) and $\mathbb{1}$ denotes the Heaviside function, which is 1 if its argument is nonnegative and 0 otherwise. To wrap up, the model development procedure is summarized in Fig 8.1.

After the parameter vector $\boldsymbol{\Theta}$ and the model order L have been estimated, we can simulate the dynamic response of the system under new test scenarios using iterative one step predictions. Since it is the first time that GPs are deployed to model ADNs, the "naive" approach was implemented, in which the input vector $\boldsymbol{x}_t$ at time step t is constructed using the normalized input and output values as:

$$\begin{aligned} \boldsymbol{x}_t = [&\Delta V_n(t), ..., \Delta V_n(t-L), \\ &\Delta \hat{Y}_n(t-1), ..., \Delta \hat{Y}_n(t-L)], \end{aligned} \tag{8.15}$$

where subscript "n" denotes the normalized signals. Despite the overconfidence characteristic of the "naive" approach, the validation results manifested the efficacy of this method in practice. To clarify the architecture of a dynamic ADN model

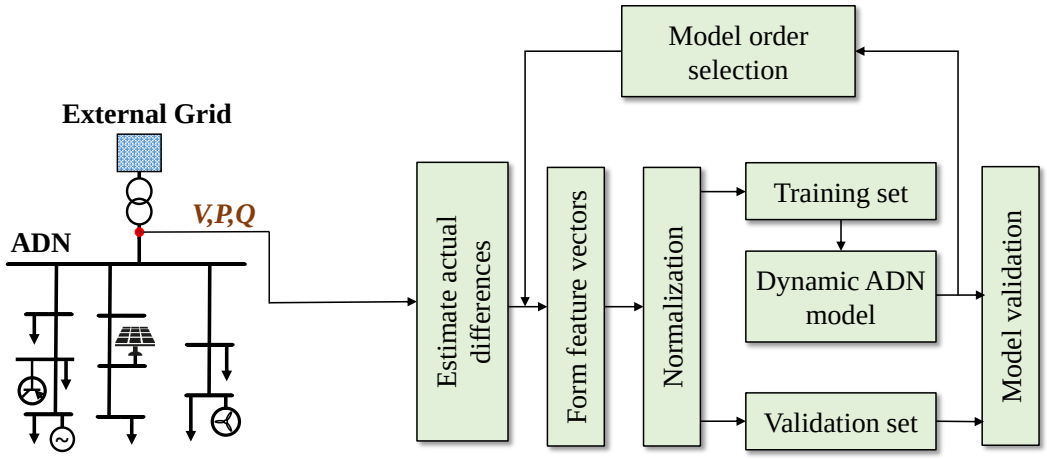


Figure 8.1: Schematic representation of the proposed model development procedure.

based on GP, the schematic representation of the proposed model structure is illustrated in Fig 8.2, where z denotes the lag operator. Finally, the ADN response can be reconstructed by simply adding the initial power Y_0 to the actual difference $\Delta\hat{Y}(t)$ as generated by the GP model and then denormalized. This can be mathematically written as:

$$\hat{Y}(t) = \Delta\hat{Y}(t) + Y_0. \tag{8.16}$$

Here, it is worth mentioning that the GP model follows the Bayesian inference;

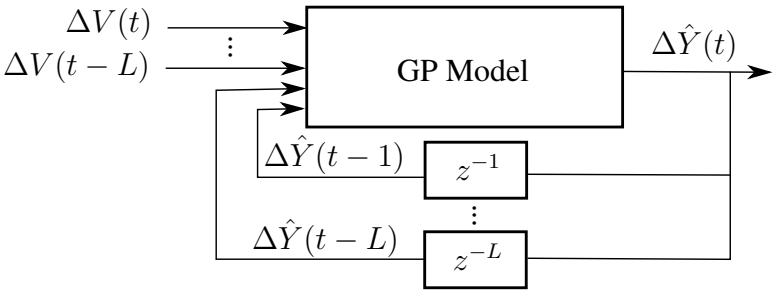


Figure 8.2: Schematic representation of the proposed GP model.

meaning that the more diverse recordings are obtained, the more informative the model will be. Hence, in case of a completely different input compared to the previously seen, the model may not be able to accurately extrapolate. Nevertheless, the variance of the prediction will be high, indicating the low confidence of the model over its output.

8.3.2 An indicative example

An indicative example is illustrated in order to emphasize the probabilistic characteristic of the GP and how the model adjusts its predictive uncertainty based on the given training data. To do so, we train a GP model based on a small random subset of 50 measurements acquired in Sub.A.1 and depicted in Fig. 8.3. The gray curves illustrate the individual measurements in terms of voltage, active, and reactive power change as calculated by (8.12), while the black bold curves denote the average of the individual measurements. It is worth reminding that the signals corresponding to voltage step-downs ($\Delta V < 0$) were multiplied by -1 only for a visually better and compact representation of the underlying dynamics.

As shown in Fig. 8.3, all the acquired data represent voltage steps resulting in active power responses following the classical load recovery behavior due to the

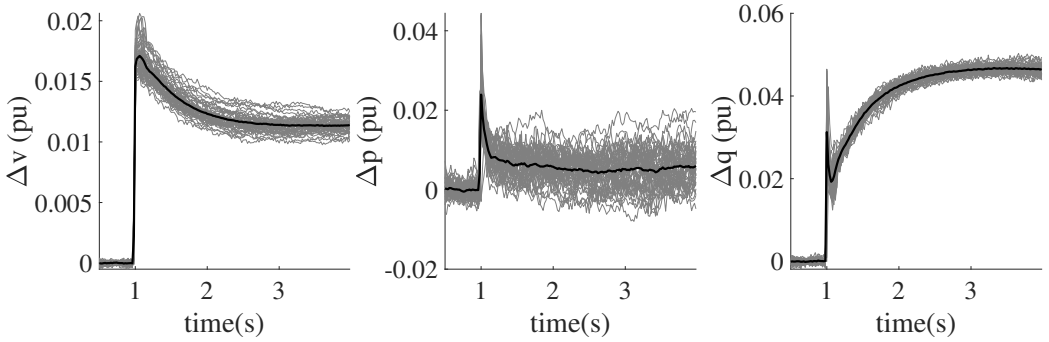
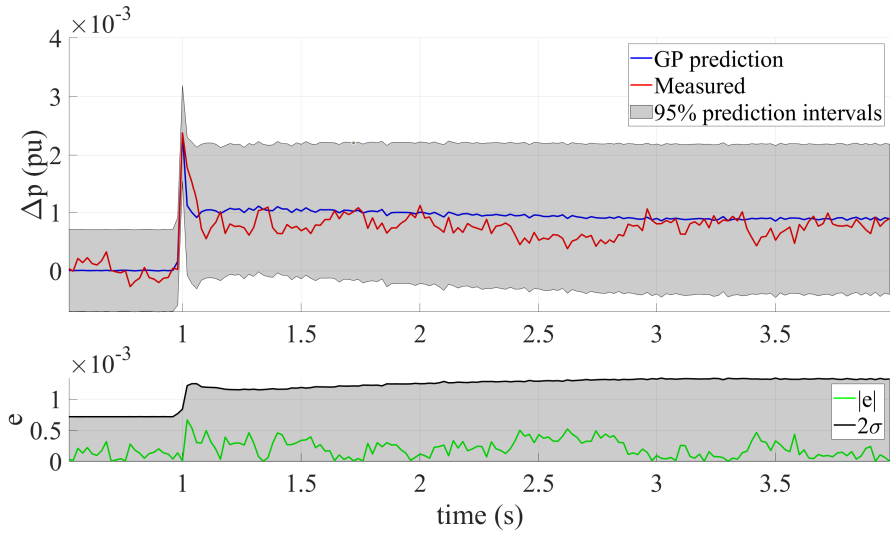
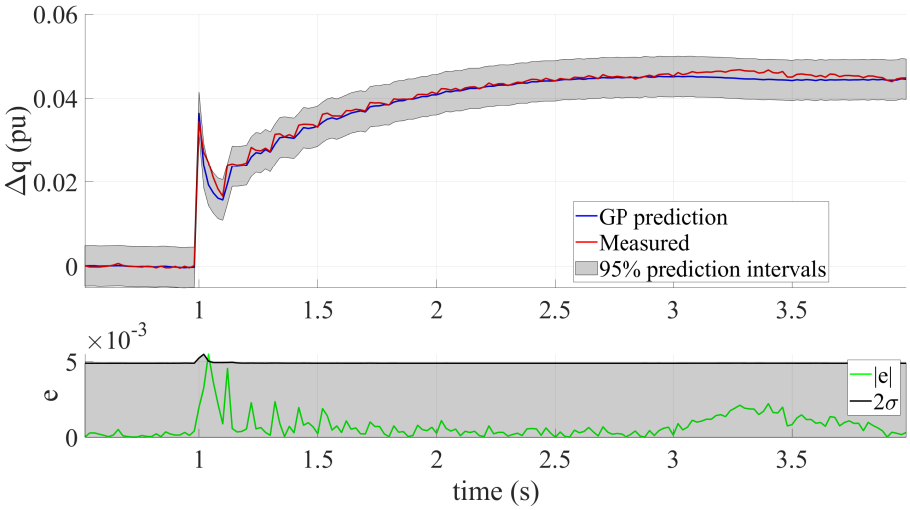


Figure 8.3: Actual differences in terms of voltage, active, and reactive power of 50 acquired measurements in Sub.A.1.



(a) High predictive uncertainty.



(b) Low predictive uncertainty.

Figure 8.4: Model prediction for active (a) and reactive power (b).

induction motors' presence within the examined system [67]. On the contrary, Δq plots show a different dynamic behavior, in which the reactive power starts increasing after the voltage disturbance. This is caused by the installed DG controllers that start providing voltage support through reactive current injections.

In Fig. 8.4, a representative response selected from the validation set is illustrated

for both active and reactive power. It can be clearly observed that both models yield dynamic responses very close to the measured values. Both predictions lie within the 95% confidence limits, as the absolute value of the prediction error e does not surpass the 2σ curve in Fig. 8.4. At the same time, the two models differ considerably in the predictive uncertainty, as shown by the gray area defined within $\pm 2\sigma$. This result originates partly from the fact that the active power measurements are characterized by higher noise levels, due to the continuous switching of the consumers. The second source of this higher uncertainty stems from the relatively big variance in the active power responses despite the similarity in the voltage disturbances. This phenomenon is caused by the different load amount and composition that vary from hour to hour and day to day affecting the dynamic properties of the system. In contrast, the variance of the reactive power measurements is quite low and thus, the confidence over the corresponding mean prediction is high. A slight increase in the prediction uncertainty is noticed at the time step of the voltage jump, due to the higher variance observed in the measured reactive power signals at that moment, as shown in Fig. 8.3. It is worth pointing out that the prediction intervals do not vary significantly over time due to the use of the "naive" approach for the multi-step prediction. In this regard, the yielded uncertainty originates from the variance of the training set induced by the stochastic nature of the load and RES.

Finally, this indicative example aims to spotlight how the stochastic nature of the load is reflected in the acquired data and consequently, how this uncertainty is captured by the proposed model and is translated into a distribution over the possible output values.

8.3.3 Results

Individual results

It has been shown in Chapter 5 that the same ADN may exhibit an array of different dynamics, which can be grouped into clusters of similar behavior such as the one shown Fig. 8.3. In this context, one to three clusters were generated in each ADN, where each cluster comprises two GP models, one for active and one for reactive

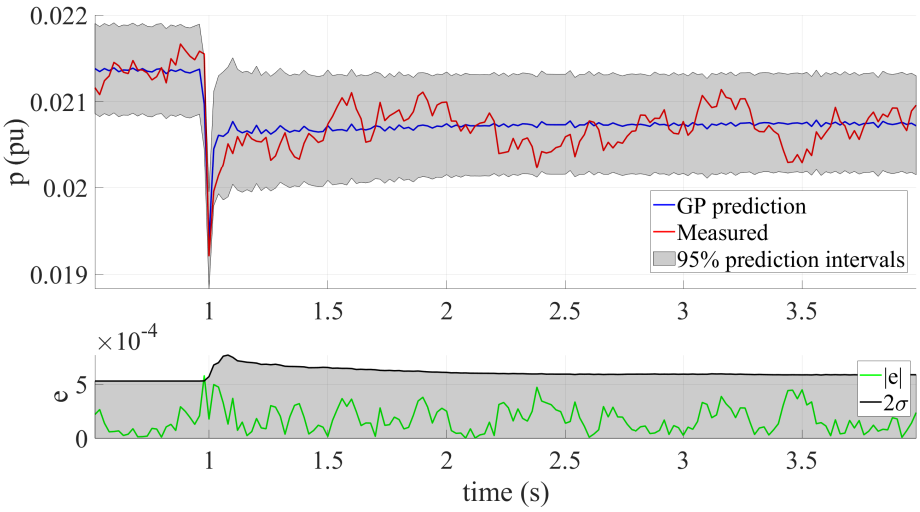


Figure 8.5: Sub.A.1: Representative active power response; prediction error and confidence plots.

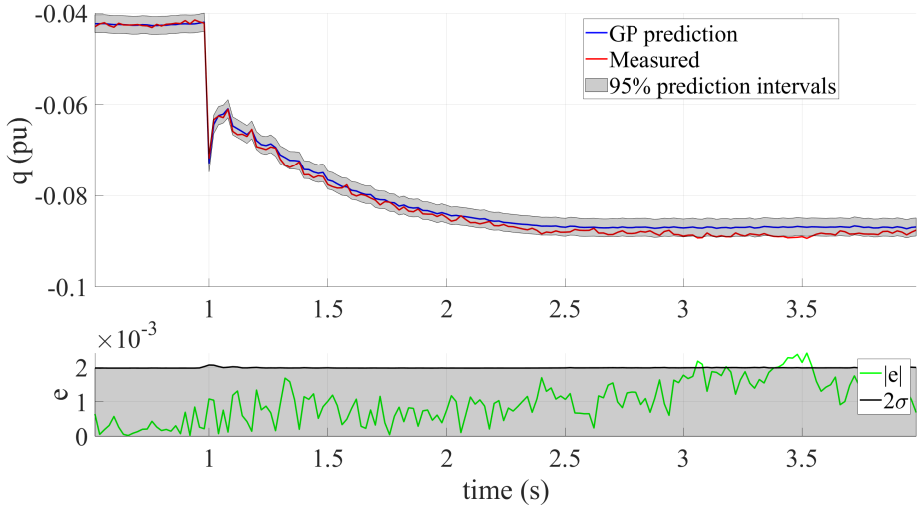


Figure 8.6: Sub.A.1: Representative reactive power response; prediction error and confidence plots.

power. In order to avoid showing several plots with identical dynamics, indicative examples of different dynamic behavior are presented in Fig. 8.5-8.9. Those fig-

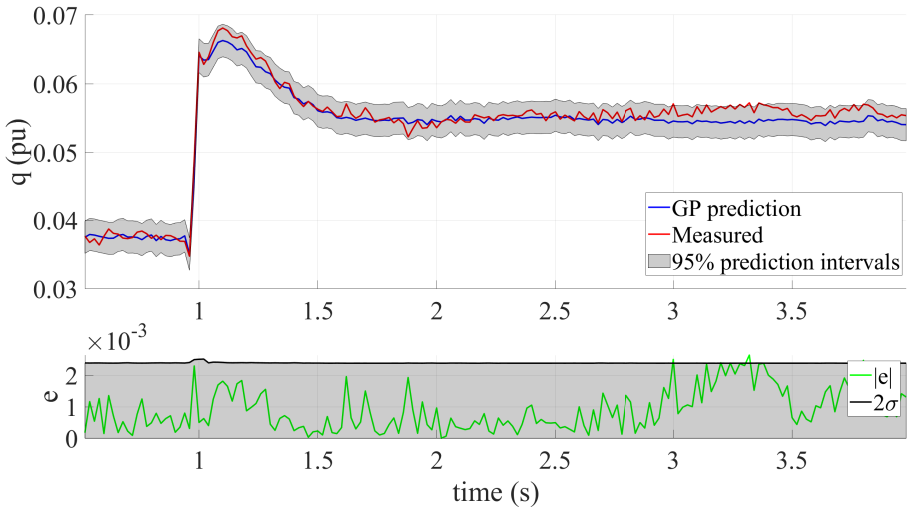


Figure 8.7: Sub.B.1: Representative reactive power response; prediction error and confidence plots.

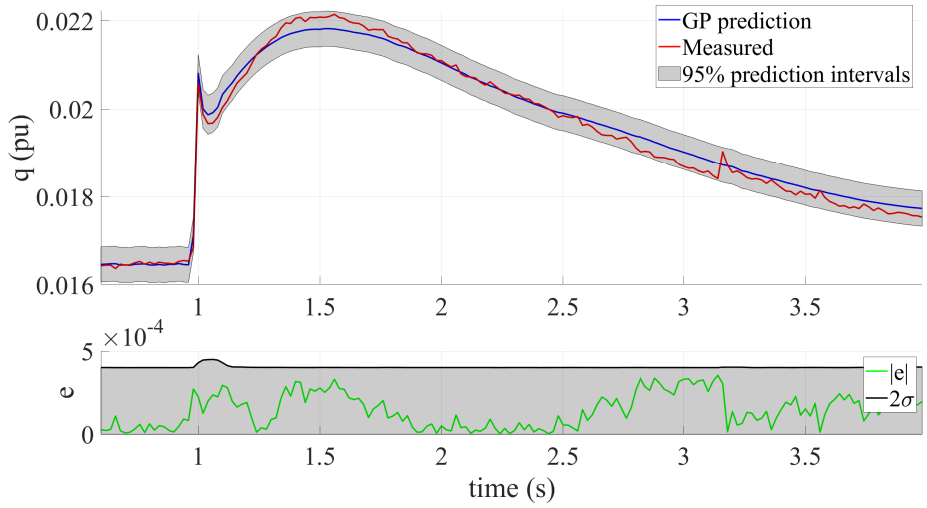


Figure 8.8: Sub.C.1: Representative reactive power response; prediction error and confidence plots.

ures demonstrate the capability of GPs to capture utterly different ADN dynamics. Fig. 8.5 correspond to active power whereas Fig. 8.6-8.9 corresponds to reactive

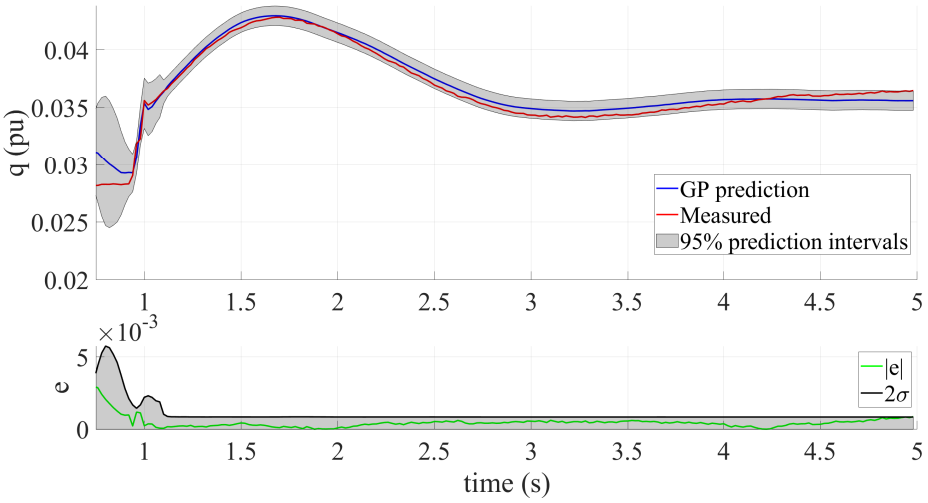


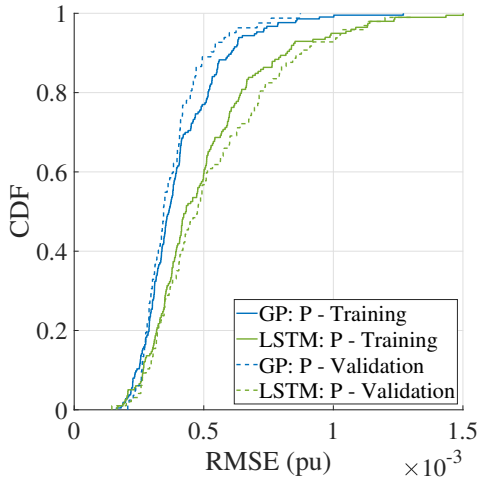
Figure 8.9: Sub.D.1: Representative reactive power response; prediction error and confidence plots.

power. Note that those measurements were selected from the respective validation sets.

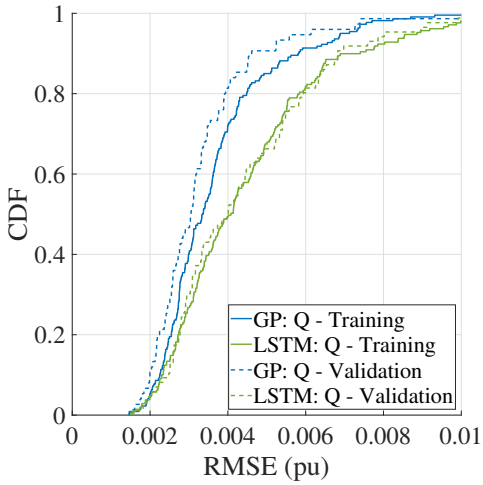
More specifically, almost all of the derived GP models require second order regressors in order to accurately capture the observed ADN dynamics. Those second order models can be categorized in two groups. The first category includes all the models with the classical load recovery behavior, where the active or reactive power immediately reacts to a voltage step and then partially recovers to a new steady-state (Fig. 8.5). The second category contains the dynamics of several voltage controllers, which support the voltage through reactive power injections (Fig. 8.6 and 8.7). Only two reactive power models of Sub.C.1 and Sub.D.1 require fifth order GP models. Those two models correspond to the slower dynamics of their voltage controller (Fig. 8.8 and 8.9) and thus, a lower order model was not able to capture that.

Comparison with a deep learning model

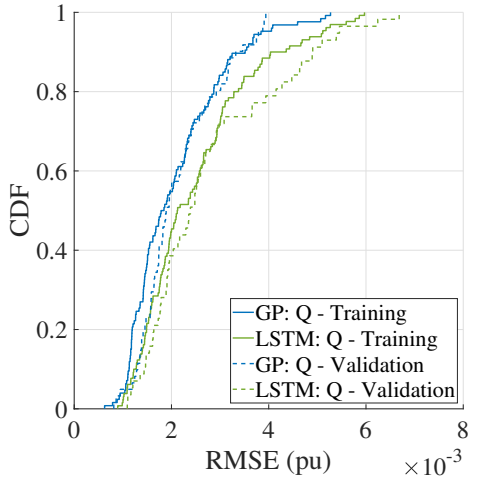
The aim of this section is to compare the proposed GP model with a state-of-the-art deep learning model, which leverages an LSTM network to capture the nonlinear



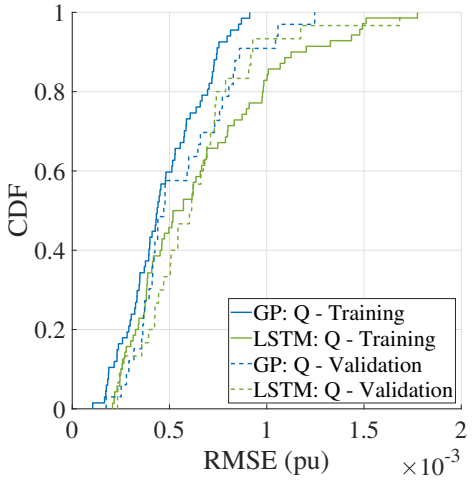
(a) Sub.A.1: Active power.



(b) Sub.A.1: Reactive power.



(c) Sub.B.1: Reactive power.



(d) Sub.C.1: Reactive power.

Figure 8.10: CDFs of the RMSEs yielded by the proposed GP model and the LSTM of [3] for different ADN dynamics.

dynamics of an ADN [3]. In our application, the same LSTM architecture and hyperparameters are used as in [3], where the network was built using 50 LSTM blocks and was trained using the Adam optimizer [135]. To thoroughly compare the proposed model with the LSTM network, we deploy the cumulative distribution

functions (CDFs) of the yielded RMSEs between the respective model output and the real values. Note that the RMSEs of the various clusters were merged into one group for every ADN and for every load/ADN model.

The CDFs of four different ADNs are exhibited in Fig. 8.10. Those ADNs are characterized by different dynamics including the classical load recovery behavior (Fig. 8.10a) and three different voltage support schemes (Fig 8.10b-8.10d). Importantly, qualitatively similar results are obtained for each of the ADNs under study. More specifically, the proposed GP model enjoys a slightly higher predictive accuracy than the respective LSTM in ADN modeling. In addition, although both approaches are characterized as black-box models, and thus there is no physical explanation over their predictions, the output distribution of the GP model could provide a meaningful insight about the time-varying dynamics of the system and to which extend they may vary. However, the training time of the derived GP models is significantly larger than the respective LSTM networks, despite the fact that the former comprises only three parameters whereas the latter around 5000.

It is also worth mentioning that training and validation sets yield identical RMSEs in both approaches; a fact that stresses their high generalization capability. However, it is of utmost importance that the GP models should be trained with a large number of diverse scenarios, since they follow the Bayesian modeling framework.

8.4 Dynamic modeling of ADNs using Gaussian processes with exogenous variables

The stochastic nature of load and DG poses many challenges in the development of realistic dynamic models for ADNs. The consumers constantly alter their power demand both in magnitude and composition while the DG output is governed by weather conditions. To this end, a dynamic ADN model is proposed in the section that can directly incorporate time and weather variables in order to decode their influence and eventually yield more accurate predictions. To do so, a GP model is

tailored such that it can capture the nonlinear ADN dynamics while learning any relevant information encoded in a list of exogenous variables such as time of the day, month, air temperature, etc.

8.4.1 Methodology

For the development of a GP model with exogenous variables, the same modeling principles as in the previous section are followed. More specifically, M voltage disturbances are acquired at the PCC corresponding to the RMS values of voltage, active, and reactive power signals. Then, the actual differences of recorded signals are calculated using 8.12.

In the next step, hourly time series data of temperature, solar radiation, wind speed, and humidity are obtained from the German Weather Agency (Deutscher Wetterdienst - DWD) [112] using the closest weather stations to the examined ADNs. Then, those weather data are interpolated in order to estimate the weather conditions at the moment of each disturbance.

Subsequently, a feature vector $\boldsymbol{x}_{j,t}$ is created for each time step t and for each measurement i as:

$$\begin{aligned} \boldsymbol{x}_{i,t} = [&\Delta V_i(t), ..., \Delta V_i(t-L), \\ &\Delta Y_i(t-1), ..., \Delta Y_i(t-L), \\ &\mathrm{v}_i^1(t), ..., \mathrm{v}_i^O(t)], \end{aligned} \tag{8.17}$$

where L indicates the model order, i.e., how many previous time steps are considered, while v^o is the o-th exogenous variable. Note that each feature vector $x_{i,t}$ corresponds to the output value of the same time step: $y_{i,t} = \Delta Y_i(t)$. The selection of the exogenous variables is based on an explanatory data analysis that reveals which of the variables contain information about the system dynamics. The steps of the proposed explanatory data analysis can be found in Chapter 3. As a next step, all individual feature vectors are combined into a feature matrix $\mathbf{X}$ and the output values are stacked into an output vector $\boldsymbol{y}$, respectively. Both of them are normalized within $[-1, 1]$ and then split into training and validation set. In our application,

70% of the data is employed in the training of the GP model whereas the rest 30% is used for validation.

To determine the model order, a same strategy as in the GP model with no exogenous variables is employed. In particular, starting from $L = 1$, we sequentially increment it by 1 while keeping track of the average CRPS over the training set. Once the difference between two consecutive CRPSs does not surpass a user-defined threshold, the training is stopped and L is set to its penultimate value. To wrap up, the model development procedure is summarized in Fig 8.11.

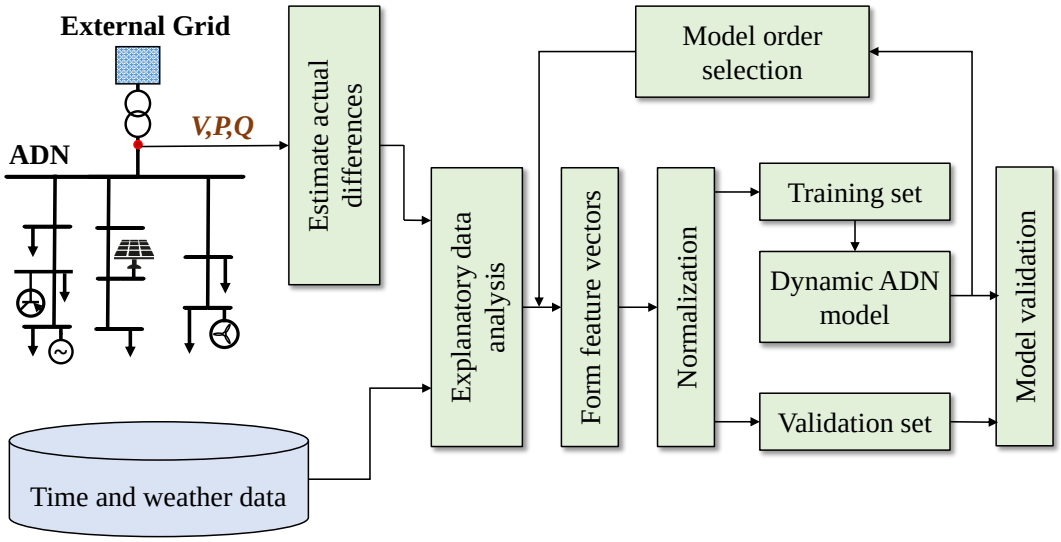


Figure 8.11: Schematic representation of the proposed model development procedure using exogenous variables.

Regarding the simulation of new test scenarios, iterative one step predictions are deployed in order to reconstruct the dynamic response of the ADN. To do so, the predicted values of the previous steps are used as features for predicting the current step. This procedure is summarized in Fig. 8.12, where z denotes the lag operator. Finally, the actual power response of the ADN $\hat{Y}(t)$ can be readily estimated by adding the denormalized model output $\Delta\hat{Y}(t)$ to the initial pre-disturbance power value Y_0 as:

$$\hat{Y}(t) = \Delta\hat{Y}(t) + Y_0. \tag{8.18}$$

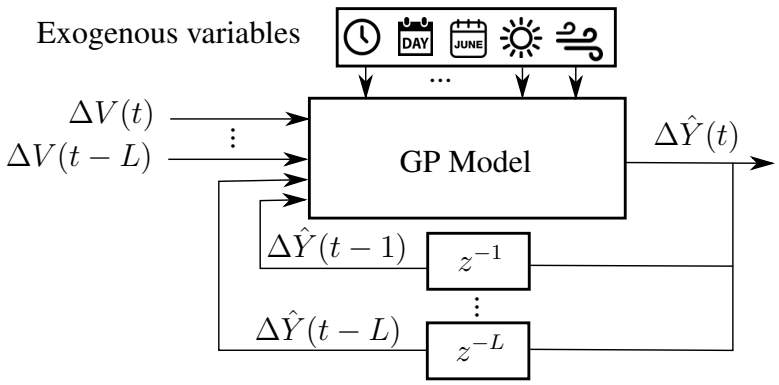


Figure 8.12: Schematic representation of the proposed GP model.

8.4.2 Results

Since probabilistic machine learning models are in principle computationally demanding, developing dynamic models for all possible combinations of exogenous variables would result in very long training times and most likely in redundant models. Hence, it is important to identify the exogenous variables that actually contain information about the load or ADN dynamics and not blindly include all of them in the model. To do so, an explanatory data analysis was performed in Chapter 3 using the exogenous variables listed in Table 8.1. Note that the temporal features are selected based on the work of [25, 26, 39], which underlines their influence on load dynamics, while the respective weather features correspond to four major factors affecting the DG [113, 114].

Table 8.1: Exogenous variable list and notation.

Temporal features		**Weather features**	
h	time of the day (hour)	t	temperature
d	day (weekday or weekend)	r	solar radiation
m	month of the year	w	wind speed
		u	humidity

Comparison of two Gaussian process models

Based on the results of the explanatory data analysis, a few features may be completely discarded for the respective ADNs. For instance, wind speed is not included in the feature list of the active power models in Sub.A.1 due to the lack of correlation, while similarly, humidity is discarded in the reactive power models of the same ADN. Then, all possible combinations of the selected features were examined resulting in numerous dynamic active and reactive power models.

Table 8.2: Validation set: Average performance metrics using the derived models with and without exogenous variables.

Sub.A.1	P	P_m	% diff.	Q	Q_dt	% diff.
RMSE (pu)	0.0037	0.0036	-2.92	0.0046	0.0043	-6.40
CRPS (pu)	0.0022	0.0021	-3.47	0.0028	0.0025	-11.44
Sub.B.2	P	P_ht	% diff.	Q	Q_hdt	% diff.
RMSE (pu)	0.0031	0.0026	-15.08	0.0032	0.0031	-3.56
CRPS (pu)	0.0019	0.0016	-16.15	0.0022	0.0021	-4.39
Sub.C.1	P	P_hd	% diff.	Q	Q_md	% diff.
RMSE (pu)	0.00036	0.00034	-4.61	0.00041	0.00036	-13.05
CRPS (pu)	0.00020	0.00020	-4.17	0.00025	0.00024	-3.38
Sub.E.1	P	P_hd	% diff	Q	Q_md	% diff
RMSE (pu)	0.0012	0.0012	-3.30	0.0012	0.0010	-15.44
CRPS (pu)	0.0007	0.0007	-5.12	0.0007	0.0006	-13.60

The best combinations in terms of accuracy and predictive uncertainty are presented in Table 8.2 for four indicative ADNs. Nevertheless, similar levels of increase in accuracy ranging from 3% to 16% were observed in each of the examined ADNs. The models denoted by a single letter, i.e., P or Q, represent the derived dynamic models using only the lagged input-output values (without considering exogenous variables) whereas the models P_x or Q_x indicate the proposed GP models

using exogenous variables for active and reactive power, respectively. The subscript 'x' corresponds to the combination of the exogenous variables based on the notation introduced in Table 8.1. Note that the accuracy of the model output is assessed using the RMSE between the predicted dynamic responses and real recorded values. As for the whole predicted distribution, the CRPS is employed, see (8.14). The presented RMSE and CRPS values correspond to the average values estimated using the validation sets.

Importantly, the dynamic models developed including a set of exogenous variables as features achieve significantly better results. In particular, the average RMSE and CRPS are improved by around 3% to 16% depending on the power model (active or reactive) and the ADN. Those results manifest that the proposed GP model structure can actually learn the important information about the system dynamics encoded in temporal and weather variables and then translate it into more accurate predictions. Interestingly, temporal variables, i.e., time of the day (hour), day of the week, and month, seem to have the greatest influence on system dynamics as each of the derived models includes at least one of them. Regarding the weather

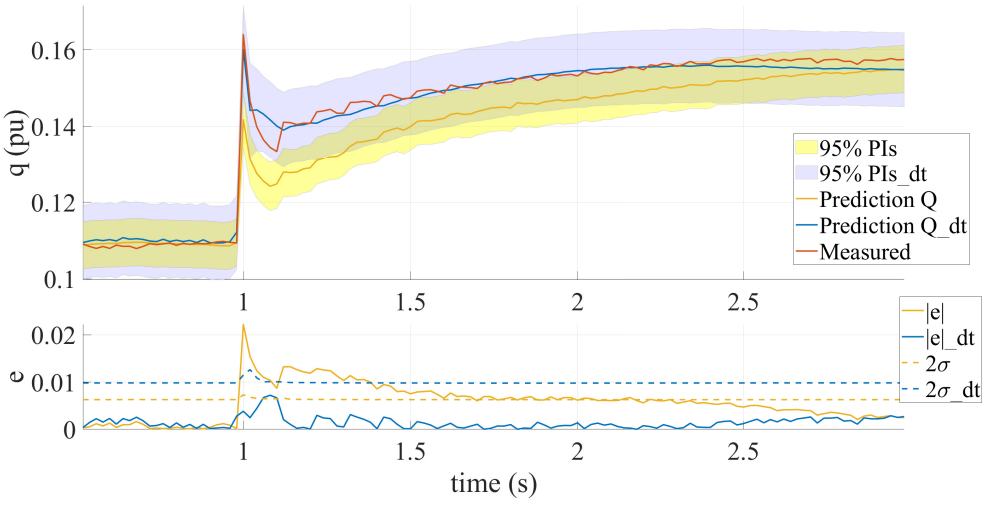


Figure 8.13: Sub.A.1: Reactive power response governed by a DG voltage controller supporting the voltage through reactive current injections.

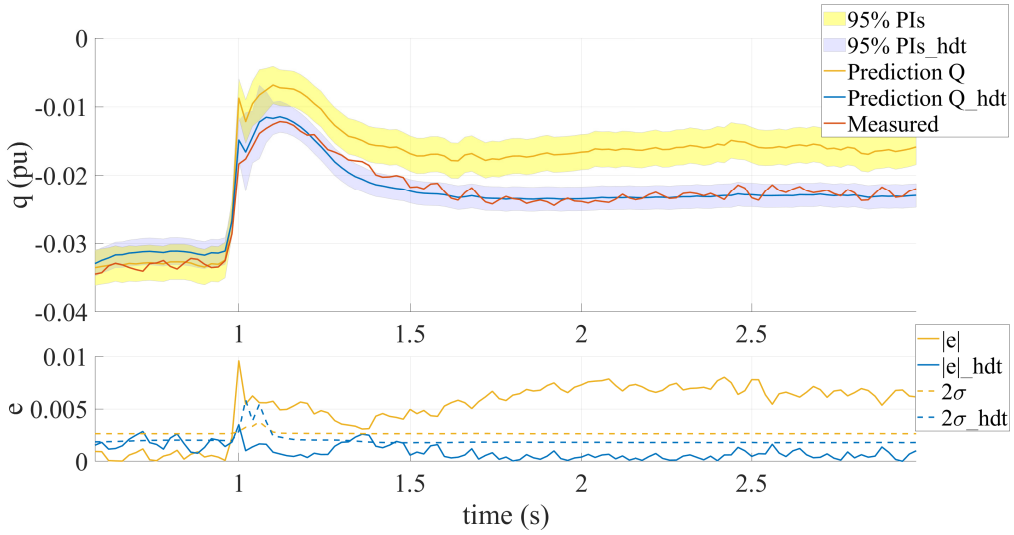


Figure 8.14: Sub.B.2: Reactive power response governed by a DG voltage controller supporting the voltage through reactive current injections.

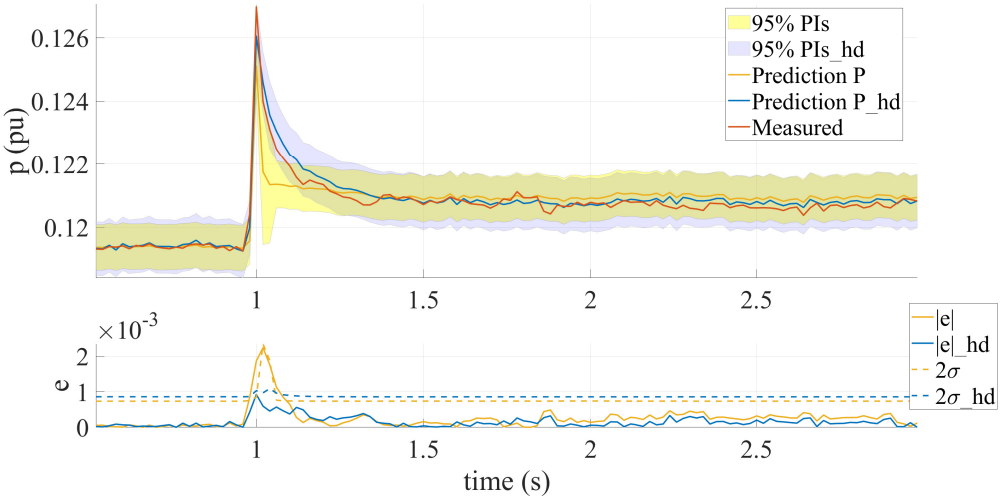


Figure 8.15: Sub.C.1: Active power response following the load recovery behavior.

variables, the temperature data can also be beneficial since around 25% of the generated models exploit this information to yield better predictions. Moreover, the deployment of all the available features did not achieve the best results in terms of

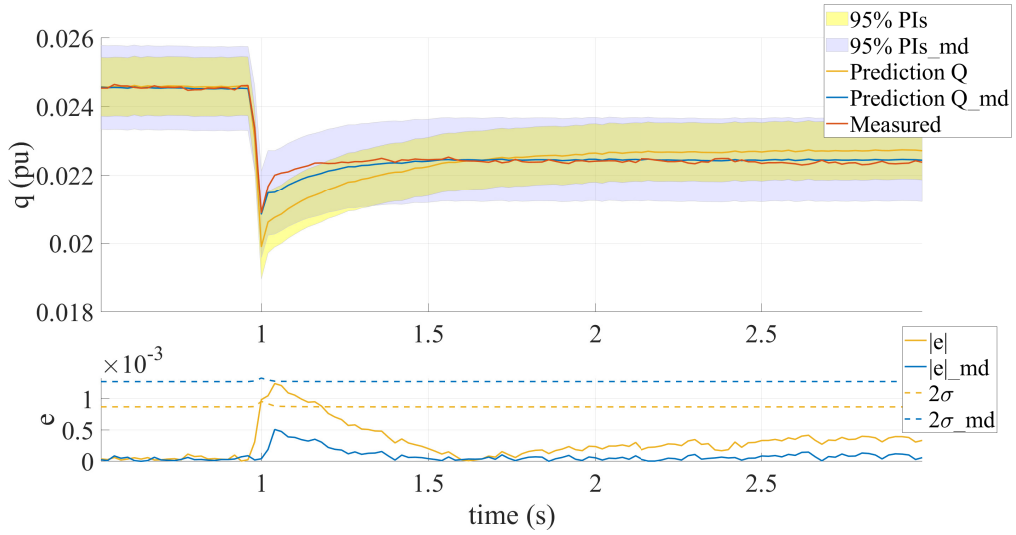


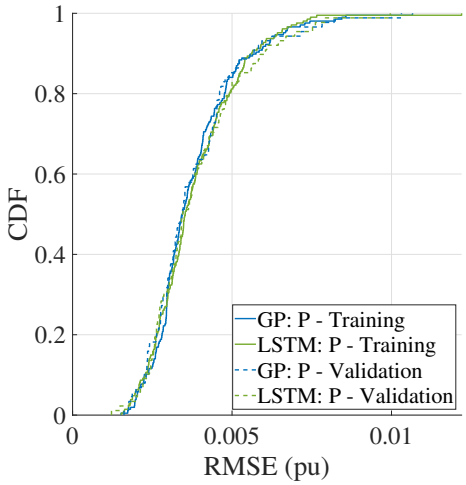
Figure 8.16: Sub.C.1: Reactive power response following the load recovery behavior.

accuracy. This may be attributed to the fact that local optimal solutions were found due to the higher dimensionality of the data and/or weakly correlated features do not actually contain useful information about the ADN dynamics.

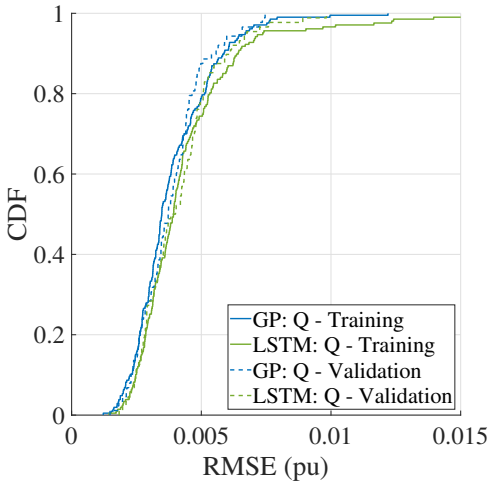
A few indicative examples randomly selected from the validation set are illustrated in Fig. 8.13-8.16 showing mainly the generated probability distributions (upper plots) as well as the prediction errors $e(t) = y(t) - \hat{y}(t)$ (lower plots). It is clearly observed that in all four examples, the dynamic models with exogenous variables yield dynamic responses very close to the real recorded values while the latter lie almost always within the 95% prediction intervals. On the contrary, the models with no exogenous variables may have a small offset in their output. This may be caused by the fact that identical voltage disturbances can lead to different ADN responses due to different load composition.

Comparison with a deep learning model

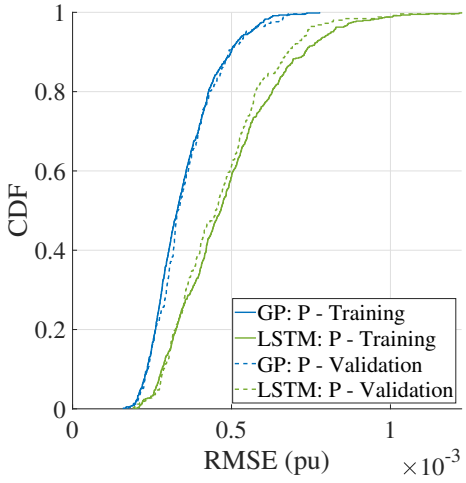
Here, we compare the proposed GP model with the deep learning model based on long short term memory (LSTM) networks of [3]. To fairly compare those two



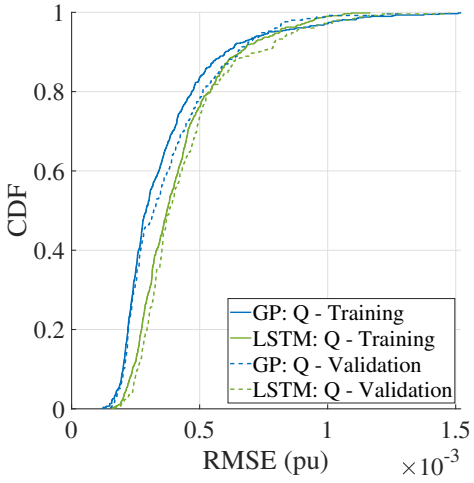
(a) Sub.A.1: Active power.



(b) Sub.A.1: Reactive power.



(c) Sub.C.1: Active power.



(d) Sub.C.2: Reactive power.

Figure 8.17: CDFs of the RMSEs yielded by the proposed GP model with exogenous variables and the LSTM of [3] for different ADN dynamics.

models, we extend the LSTM network so that it additionally includes the respective exogenous variables. The network architecture as well as the hyperparameters remain the same as in [3], where 50 LSTM blocks are employed and the training was performed using the Adam optimizer [135]. Moreover, the CDFs of the RMSEs

between the respective model output and the real recorded signals are used in order to rigorously compare the accuracy of each model. Note that the RMSEs of the various clusters were merged into one group for every ADN and for every load/ADN model.

The CDFs of two different ADNs are exhibited in Fig. 8.17 for both active and reactive power. Those ADN models are characterized by different dynamics and different combinations of exogenous variables. As depicted in Fig. 8.17, the proposed GP model enjoys a slightly higher predictive accuracy than the respective LSTM. Importantly, qualitatively similar results are obtained for each of the ADNs under study, where both approaches yielded almost identical CDFs. Nevertheless, the LSTM method does not provide any confidence about its predictions whereas the proposed GP approach yields a probability distribution revealing also possible trajectories of the dynamic response as well as borderline scenarios. However, the probabilistic nature of the proposed GP model requires larger training times than the respective LSTM despite the small number of model parameters that need to be estimated.

8.5 Summary

The development of an accurate yet generic dynamic equivalent model for active distribution networks (ADNs) is a rather delicate task due to the high uncertainty characterizing the various grid components. To this end, a probabilistic dynamic model was proposed that, except for its response, additionally yields the corresponding predictive uncertainty. The proposed model learns the underlying dynamics through measurement data and yields Gaussian distributions as output.

This probabilistic output is one of the major differences between the existing ADN models and the proposed one. Importantly, the experimental results confirmed the applicability of the proposed model in practice and how load and DG uncertainties can be reflected in the model output. Therefore, the proposed probabilistic model can serve as a valuable tool in order to quantify uncertainty and

additionally deliver confidence intervals alongside with its predictions.

Furthermore, the experimental results indicate that the RBF kernel can be a reliable choice for a covariance function. More complex kernels may simply increase the computational need without necessarily offering a significant increase in the predictive accuracy. Importantly, the choice of a kernel adds prior knowledge about the system to the model. For instance, if there are oscillations in the output, a reasonable approach would be to use the periodic kernel. Therefore, the choice of the kernel highly depends on the characteristics of the measurement data.

Apart from the kernel function, a very crucial factor for developing reliable GP models is the data themselves. For instance, there is a great need for diversity in the measurement data deployed in the training. At the same time, more data means more resources in terms of computation. Hence, efficient training techniques should be also considered in order for the GP models to incorporate more knowledge about the system. In the proposed method, a simple data sampling technique was followed. In addition, low order models seem to suffice for modeling ADN dynamics, but no general conclusion could be extracted about the model order nonetheless.

Based on the model validation, the use of the "naive" approach in multi-step predictions proved to be a reliable choice to accurately capture the ADN dynamics. Nevertheless, the implementation of a multi-step prediction using the predictive uncertainty of the previous steps as well as a thorough comparison of the two approaches ("naive" and "exact") are yet to be explored and are suggested as future work.

In the second part of this chapter, the proposed dynamic GP models were further extended. A new model structure was introduced that can reflect the time-varying and weather dependent dynamics of load and DG by incorporating a set of exogenous variables, such as time of the day, day of the week, month, temperature, solar radiation, and wind speed. The proposed model can learn the influence of this kind of exogenous variables and convert it into more accurate predictions. In particular, the proposed method was applied in the challenging problem of dynamic ADN modeling. It is shown that, among a list of exogenous variables, the temporal ones can be the most advantageous since they can encode important information about

the system dynamics. However, no conclusion was drawn about the type of ADN and the list of exogenous variables yielding the best results. That would probably require data from significantly more substations.

Moreover, comparative results indicated that there may be an offset in the generated dynamic responses, when no exogenous variable is used, especially during the transient phase. Although this may seem negligible for a single ADN model, power system models are composed by several of those single load/ADN models. As a result, an aggregation of those offsets might lead to false conclusions about system stability. By incorporating exogenous variables in the ADN models, the overall simulation accuracy is increased and thus, the risk of false conclusions in the context of stability studies is decreased.

The comparison with the LSTM approach reveals that, apart from the proposed GP model, there are also other model structures that can be modified accordingly in order to decode the influence of temporal and weather features and integrate them into their predictions. Hopefully, this outcome will motivate more studies in power system dynamics to incorporate that kind of information.

Finally, it is worth discussing the applicability of those probabilistic models in power system dynamic simulations. Although those models offer the advantages discussed earlier in this chapter, there is no commercial software, at the time of writing this thesis, that can perform probabilistic dynamic simulations. The following chapter reviews how those models can be integrated into a power system simulation for stability studies.

Chapter 9

Deployment of the proposed models in stability studies

This chapter starts with a description of the international industry practices on power system load modeling. Next, this chapter discusses the applicability of the proposed models and gives guidelines for their optimal integration in power system simulation packages.

9.1 Motivation

In the past, TSOs traditionally used static models to simulate the behavior of loads in power system simulations [48, 144]. Those models describe the relationship between power (active or reactive) and voltage and frequency using algebraic equations. Therefore, they are particularly suitable for steady-state analysis, e.g., power flow estimation [144]. Nevertheless, those models may not be able to deliver accurate results in case of dynamic simulations, as they comprise only algebraic equations [20, 42, 145, 146]. In spite of their limitations in dynamic simulations, around 70% of the TSOs worldwide keep using static models for stability studies [17, 147]. More specifically, around 53% of the TSOs deploy the static exponential model, while around 19% use the ZIP model [144]. This fact can be attributed to three

main factors. The first one is simplicity of their model structure, the second one is that those models are available in every commercial power system simulation software while the third one is the data availability to parameterize the models [17, 147]. Under those conditions, it is of utmost importance to provide dynamic load/ADN models that can be easily implemented or integrated in power system simulation packages. In the next section, guidelines are provided so that the proposed models can be deployed in stability studies.

9.2 Guidelines for applying the proposed models in practice

In the next three sections, the applicability of the three proposed models is discussed, respectively. In addition, guidelines are given with respect to their practical implementations in power system simulations.

9.2.1 Linear dynamic model

As thoroughly presented in Chapter 6, the proposed active power model comprises only a linear transfer function, whereas the reactive power model is composed by two transfer functions; one for load and one for generation. For the sake of completeness, the block diagram of the proposed linear model structure is illustrated once more in Fig. 9.1.

As observed in Fig. 9.1, the block diagram comprises simple elements and calculations. Therefore, it can be readily implemented in power system simulations packages that allow the development of user-defined dynamic models. Indicative examples of popular software offering this feature are DIgSILENT PowerFactory [148], NEPLAN [149], and Matlab Simulink [150]. To further support this claim, the proposed linear model was implemented in DIgSILENT PowerFactory by a former colleague for the needs of a research project.

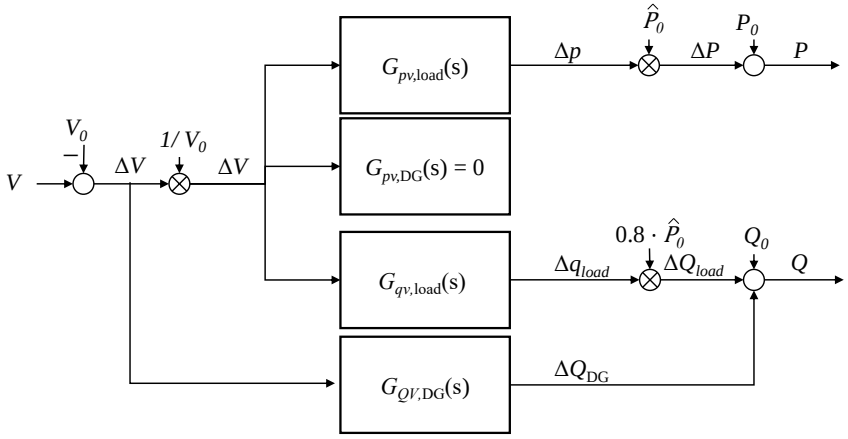


Figure 9.1: Block diagram representation of the proposed linear model structure. In case of distribution grids in which no DG with reactive power control is present, $G_{QV,\mathrm{DG}}(s) = 0$.

9.2.2 Nonlinear dynamic model based on ERM

The proposed nonlinear dynamic model follows the same fundamental model structure as the well-established ERM. In particular, it is composed by two nonlinear algebraic equations modeling the steady-state and transient response, respectively, and a linear transfer function modeling the load recovery behavior after a voltage step change. The block diagram of the proposed model structure is depicted once more in Fig. 9.2.

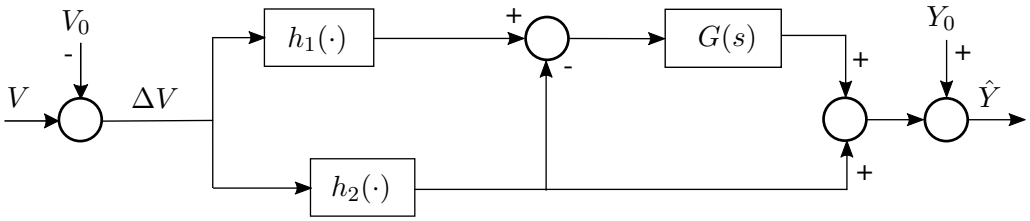


Figure 9.2: Block representation of the proposed nonlinear dynamic model.

Similarly to the aforementioned linear model, the proposed ERM structure for modeling ADNs comprises simple blocks and calculations. At the same time,

the classical ERM is one of most commonly used dynamic model structures for load modeling [18]. Hence, the proposed ERM structure can be effortlessly integrated into any power system simulation software that allow the development of user-defined dynamic models, e.g., DIgSILENT PowerFactory, NEPLAN, Matlab Simulink.

9.2.3 GP model

In case of the proposed GP model, its integration into power system simulation packages is not straightforward. The major characteristic differentiating the GP model from the rest is its probabilistic output. The GP model yields the mean and the standard deviation of a normal distribution at each time step.

At the moment that this thesis is written, there is no tool supporting any kind of probabilistic power system dynamic simulations. Therefore, the focus is placed on the generation of point predictions based on the predicted probability distributions. To do so, at each time step, the model takes as input a vector comprising the current and previous voltage values as well as the mean of the predicted distributions of the previous time steps (until a certain point as defined by the model order). Next, the model yields the mean and the standard deviation of a probability distribution and the former is selected as the point prediction. This procedure is schematically described in Fig. 9.3. The detailed description of the GP model development can be found in Chapter 8.

In order to estimate the mean of a probability distribution for each time step, the

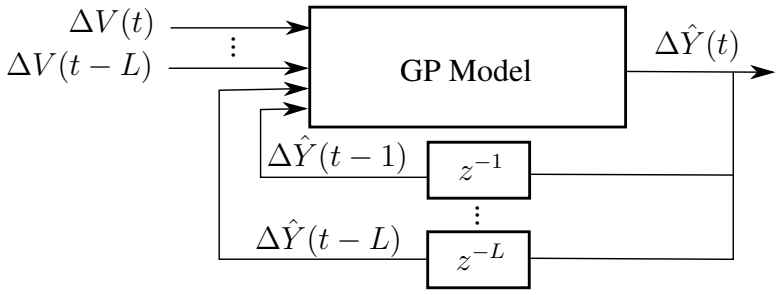


Figure 9.3: Schematic representation of the proposed GP model.

GP model block of Fig. 9.3 executes the following matrix multiplication:

$$\mu(y^*) = \mathbf{k}(\boldsymbol{x}^*)^\top \mathbf{K}^{-1} \boldsymbol{y}, \tag{9.1}$$

where $\mu(y^*) = \Delta\hat{Y}(t)$, $\boldsymbol{x}^* = [\Delta V(t), ..., \Delta V(t-L), \Delta\hat{Y}(t-1), ..., \Delta\hat{Y}(t-L)]$, $\mathbf{k}(\boldsymbol{x}^*)$ is a vector containing the covariances between each of the training input points and the test input point $\boldsymbol{x}^*$ as calculated using the respective kernel function, $\mathbf{K}$ is the covariance matrix of all the training points, and $\boldsymbol{y}$ is the vector containing all the output points used for training. In this regard, $\mathbf{K}$ and $\boldsymbol{y}$ are constants and are known in advance whereas $\mathbf{k}(\boldsymbol{x}^*)$ is estimated at each time step.

As expected, the computational requirements of the GP model rise as the number of training points increases, since the size of $\mathbf{K}$ and $\boldsymbol{y}$ increases. Nevertheless, matrix multiplication is highly optimized in most of the programming languages and software packages. In case of very large training sets, random batches of data can be used to decrease the computational requirements.

Based on the aforementioned analysis, the GP model structure (Fig. 9.3) can be integrated into any power system simulation software offering the creation of a user-defined dynamic model, where matrix multiplications can be implemented. An indicative examples of such a tool is Matlab Simulink. DIgSILENT PowerFactory plans also to incorporate matrix multiplications into its user-defined dynamic model in a future software release, as communicated by one of its core developers. It is worth mentioning that the same procedure is followed for the GP model with exogenous variables. The only difference lies in the total number of model inputs.

Incorporating uncertainty

To estimate the dynamic ADN response, the methodology described in this section considers only the mean of the predicted distribution, while the standard deviation is not used. This approach is commonly known as the "naive" approach [139]. In order to incorporate uncertainty into the predicted dynamic response, one more step is added to the "naive" approach. Instead of using the mean of the predicted distribution, the model output is sampled from it. Thus, $\Delta\hat{Y}(t)$ is a random sample from $\mathcal{N}(\mu(y^*), \upsilon^2(y^*))$, where $\upsilon(y^*)$ is the predicted standard deviation.

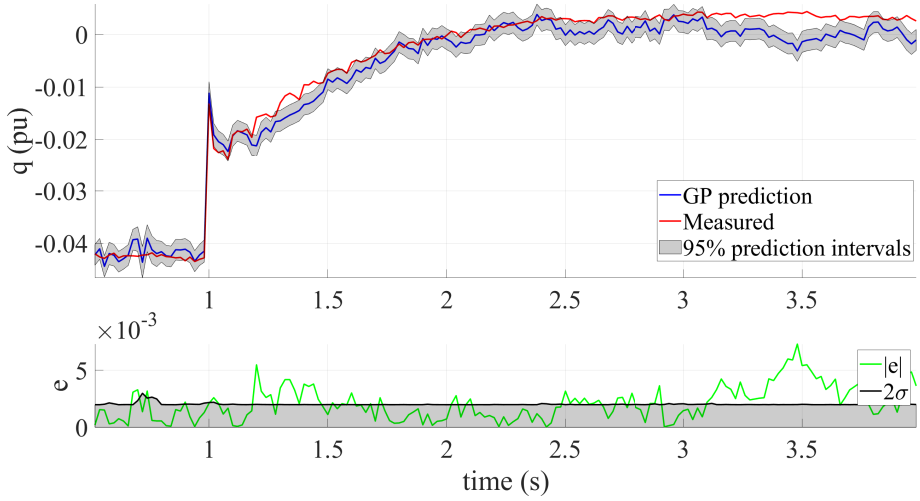


Figure 9.4: First run: Reactive power response using sampling from the predicted probability distribution.

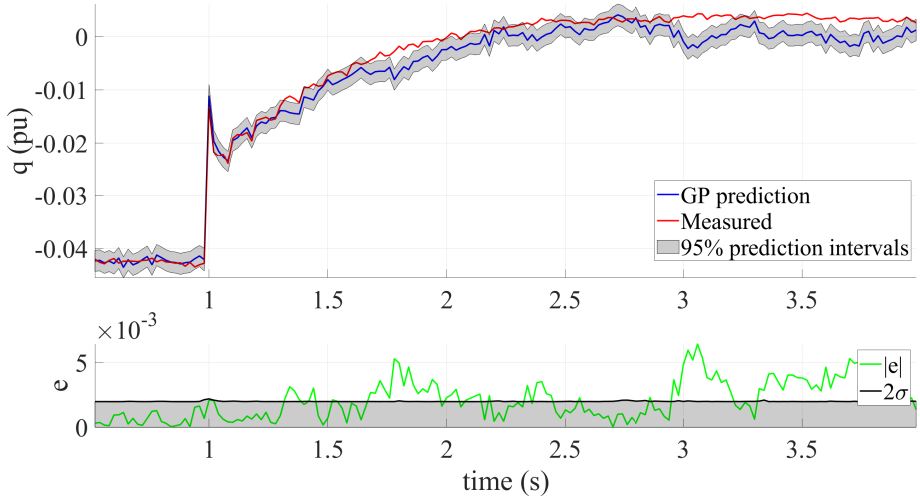


Figure 9.5: Second run: Reactive power response using sampling from the predicted probability distribution.

However, this approach yields more noisy and less accurate results. Yet each time the same power system simulation model runs, the GP model will generate dif-

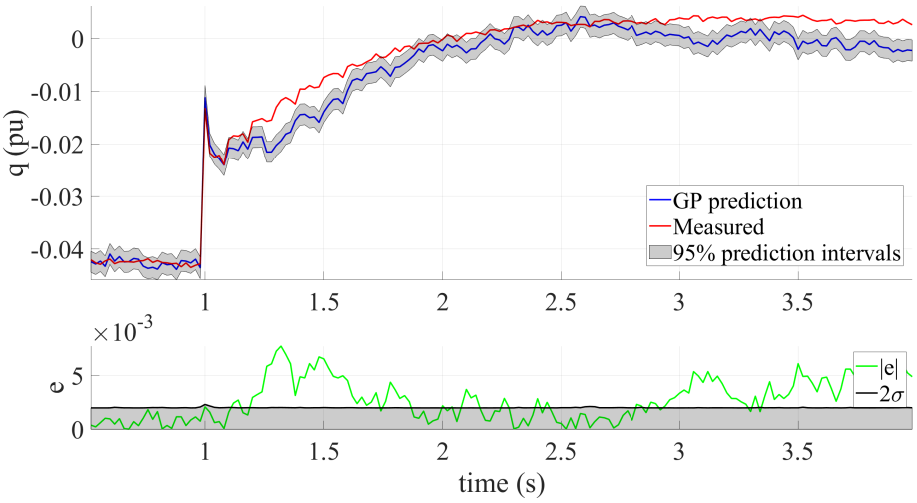


Figure 9.6: Third run: Reactive power response using sampling from the predicted probability distribution.

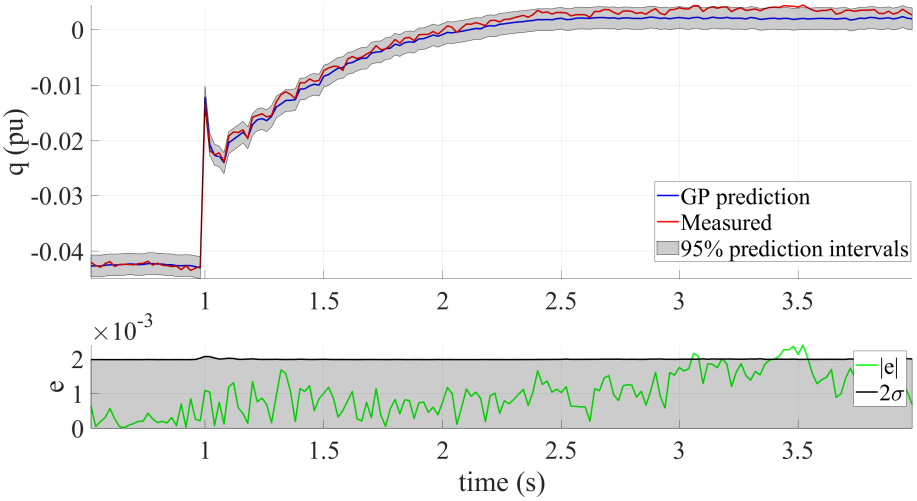


Figure 9.7: Naive approach: Reactive power response using the "naive" approach.

ferent dynamic responses highlighting its stochastic nature. For instance, Fig. 9.4-9.6 show the GP model output for the same input in three different runs. For comparison, the respective model output using the "naive" approach is shown in Fig. 9.7.

As readily inferred from those figures, the sampling approach generates noisy dynamic responses, which may deviate from the real measured values. Nevertheless, this sampling approach might be considered a way to reflect the stochastic ADN nature in stability studies. As an alternative, one could use the naive approach to generate a dynamic response such as the one in Fig. 9.7 and then, select a trajectory corresponding to a desired percentile, e.g., 75th percentile. If the desired percentile is randomly selected by a Gaussian distribution for each ADN, each ADN will generate a different dynamic response accounting for the stochastic nature of load and DG.

9.3 Summary

The motivation of developing dynamic load/ADN models is to substitute the various distribution networks in a power system model with equivalent dynamic models. This chapter provides the guidelines for integrating the proposed dynamic models into power system simulation packages. All proposed models can be implemented in a simulation software that offers the development of user-defined dynamic models as a feature. Indicative examples of that kind of software tools are DIgSILENT PowerFactory, NEPLAN, and Matlab Simulink.

In particular, the implementation of the proposed linear and nonlinear models is rather straightforward, as both models are composed by simple block elements. On the contrary, the deployment of the GP model structure may be confusing. To clarify that, the corresponding implementation details are given in this chapter. Finally, a sampling approach is proposed in order to incorporate the ADN uncertainty into the generated dynamic responses.

Chapter 10

Conclusion

10.1 Summary of this work

The ongoing transition of traditional passive load-dominated distribution systems into ADNs poses new challenges for dynamic simulations and stability studies of power systems. Hence, the applicability of the existing load models should be scrutinized while new model structures incorporating the influence of new system components, such as DG and controllable loads, should be introduced [2].

As a general rule, the measurement-based approach has been extensively deployed in the literature for load and, recently, for ADN modeling. In this approach, the derived models are developed by fitting a set of parameters to a set of input-output measurements using system identification techniques [25]. Despite the large number of approaches following the measurement-based approach that have been proposed within the past years, there are still several issues and open research questions that should be resolved.

In this dissertation, an end-to-end measurement-based approach is proposed aiming at developing dynamic ADN models while considering the current issues and limitations of the existing approaches. In particular, three EHV/HV and six HV/MV substations were equipped with measurement units, each of them acquiring data over about one year. By measuring in several substations, various grid configu-

rations were captured, while the long measuring time period allowed covering a sufficient number of events at different time and weather conditions.

Based on the acquired data, the influence of the various DG technologies on the dynamics of the systems is examined. It is concluded that the active power is not affected by any of the various DG technologies in the ADNs under study. However, this is not a general conclusion, since different control schemes may be implemented in other DG units installed in other ADNs. On the contrary, it is observed that the reactive power dynamics are highly affected by the underlying wind parks and CHP plants whereas the respective PV and Biogas plants do not have a direct impact on the reactive power responses. This conclusion cannot be generalized either and is valid only in the examined grids, as different ADNs may contain DG with completely different control schemes.

Similarly, a thorough analysis with respect to the relationship between ADN dynamics and a list of temporal as well as weather variables is performed. It was observed that the ADN dynamics may vary from hour to hour, day to day, and month to month (temporal variables). At the same time, weather variables, such as temperature, radiation, and wind speed, might have a minor yet worth considering influence on the system dynamics. Nevertheless, different relationships and correlations were discover at each ADN and thus, no general conclusion can be extracted.

Regarding the self-regulating effect, it turned out that the recorded frequency events are not suitable for the identification of the stationary relationship between the frequency and active power. As a fallback solution, two different approaches were attempted. The first approach is based on frequency changes that occurred at random moments. The second one relies on the deterministic frequency deviations happening every quarter hour due to market product design. However, no unambiguous self-regulating effect can be extracted in either of those approaches.

In the context of dynamic ADN modeling, the state-of-the-art approaches assume that the data used for the system identification are homogeneous with respect to the dynamics, neglecting the dispatch of DG units, whose operation varies based on the weather conditions. To address this issue, a novel unsupervised learning methodol-

ogy for clustering different dynamic behavior is introduced. Importantly, no knowledge of load and generation mix is required and the clustering is performed exclusively based on the measurements themselves. Furthermore, the proposed method can be easily integrated as a preprocessing step into most of the existing ADN modeling approaches. Therefore, different dynamic models can be accurately generated representing the same ADN under various operating conditions.

To evaluate how distribution networks have changed within the past years, linear load models are developed for each ADN for active power and reactive power. By combining the individual models, two generic dynamic models are derived (one for active and one for reactive power), which can be readily integrated into power system dynamic simulations for stability studies. Both models show only minor differences in the transient and steady-state responses compared to the respective average models developed several years ago.

Next, a three-stage methodology is introduced in order to effectively manage field measurements, identify the dominant dynamics, and build an adequate number of different ADN models accurately reflecting the time-varying nature of load and DG. To do so, this methodology leverages the proposed clustering method in the first two stages and a new nonlinear model structure in the third stage, respectively. This nonlinear dynamic model borrows the main concept of the exponential recovery model and is further modified in order to model modern distribution networks with significant DG. Finally, it is manifested that, in spite of the large number of measurements representing a wide range of grid configurations, the general dynamic characteristics of an ADN can be accurately captured and modeled using a limited number of clusters.

To deal with the stochastic nature of load and DG, a new probabilistic model structure based on GP is proposed. Contrary to most of the existing deterministic approaches generating single points as output, the proposed GP model yields a probability distribution at each time step. Due to its flexibility with respect to the input features, the model directly incorporates time and weather variables in order to decode their influence and eventually yield more accurate predictions. It is shown that, among a list of exogenous variables, the temporal ones can be the

most advantageous since they can encode important information about the system dynamics. However, no conclusion was drawn about the type of ADN and the list of exogenous variables yielding the best results.

This dissertation is finalized by providing the guidelines for deploying the proposed models in power system simulation packages. Indicative examples of software tools that could integrate those models are DIgSILENT PowerFactory, NEPLAN, and Matlab Simulink.

It is worth mentioning that several publications were generated based on the methods and the data presented in this dissertation. For more details, the interested reader is referred to [12, 41, 115, 116, 136, 137, 151, 152].

10.2 Directions for future work

Even though this dissertation proposes a complete methodology for developing dynamic load/ADN models including methods for exploratory data analysis, preprocessing, outlier detection, clustering, and new model structures, there are several challenges that are yet to be addressed.

Measurement data

- Although this study employs measurement data from six different HV/MV substations, which is considered a large number compared to the existing literature, measurement data from more substations spread out in the country could possibly reveal different dynamic behavior. Those field data would be highly beneficial in developing generic ADN models.

- The estimation of the steady-state relationship between frequency and active power is a rather challenging task, since big frequency changes are rare nowadays. Therefore, measuring frequency and active power over a long time period and for several ADNs (in order to improve the signal-to-noise ratio) should be pursued in order to be able to capture those rare big frequency events that are required.

Exploratory data analysis

- This dissertation presented one the first attempts found in the literature to identify correlations between load/ADN dynamics and time and weather variables. To do so, an exploratory data analysis based on correlation coefficients, boxplots, and scatter plots was proposed. This approach may be further enhanced by applying more advanced statistical analysis techniques.

Modeling

- Different types of distribution networks should be examined. The increasing interactions of the electric power system with the sectors of transportation, heat, gas, and hydrogen would introduce new load types such as, a large number of EV chargers, heat pumps, electrolyzers, etc. Thus, new model structures may be required to capture the dynamics of those grid components.

Uncertainty estimation

- To the best of the author's knowledge, this dissertation presents for the first time a probabilistic model structure that yields a full probability distribution instead of a single deterministic point. At this point, the generated prediction intervals are used in order to reason about system uncertainties inherent to measurement data, identify worst case scenarios using the boundaries of the yielded distributions, and indicate whether a prediction should be trusted, further examined or even rejected. Nevertheless, the probabilistic nature of the model output opens many opportunities for further research. For instance, those models could motivate the development of probabilistic dynamic simulations, which, to date, are not possible. However, as research in the field advances, e.g., quantum computing [153], deep learning [154, 155, 156, 157], probabilistic dynamic simulations may not be that far. To support this claim, probabilistic optimal power flow is already being studied [158, 159, 160, 161].

- Currently, the proposed model employs the "naive" approach to generate the dynamic ADN response, as it demonstrated highly accurate results. However, future work could be the implementation of a multi-step prediction by incorporating the predictive uncertainty of the previous steps as well as a thorough comparison of the two approaches ("naive" and "exact") with respect to modeling ADNs.

- It is worth pointing out that there is some recent research work that introduces stochastic differential-algebraic equations to model correlated stochastic processes in power systems [162, 163]. In this regard, the integration of the proposed probabilistic ADN models into the power system simulation environment of [162] might be a good solution to fully leverage the probabilistic nature of the models.

Large-scale studies

- This dissertation has focused on the development of generic and reliable dynamic load/ADN models that replace the individual distribution systems in the power system model. The applicability of the proposed models can be further validated by their integration in a detailed transmission system model. To date, there are limited studies dealing with this topic, as it requires reliable equivalent load/ADN models, the development of which is by itself a rather challenging task. Nevertheless, the recent work of [164] is suggested as a starting point.

Online parameter estimation

- As highlighted in this thesis, load composition and DG dispatch are time-varying and thus, different models should be used based on that. Future work could aim at the online update of dynamic load/ADN models using real-time data.

Bibliography

[1] A. Zare *et al.*, "Multistage expansion planning of active distribution systems: Towards network integration of distributed energy resources," Ph.D. dissertation, University of Saskatchewan, 2018.

[2] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load modeling—A review," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5986–5999, 2017.

[3] C. Zheng, S. Wang, Y. Liu, C. Liu, W. Xie, C. Fang, and S. Liu, "A novel equivalent model of active distribution networks based on LSTM," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2611–2624, 2019.

[4] E. O. Kontis, T. A. Papadopoulos, A. I. Chrysochos, and G. K. Papagiannis, "On the applicability of exponential recovery models for the simulation of active distribution networks," *IEEE Transactions on Power Delivery*, vol. 33, no. 6, pp. 3220–3222, 2017.

[5] N. Hatziargyriou, J. Milanović, C. Rahmann, V. Ajjarapu, C. Cañizares, I. Erlich, D. Hill, I. Hiskens, I. Kamwa, B. Pal *et al.*, "Stability definitions and characterization of dynamic behavior in systems with high penetration of power electronic interfaced technologies," *IEEE PES Technical Report PES-TR77*, 2020.

[6] H. E. Murdock, D. Gibb, T. André, J. L. Sawin, A. Brown, L. Ranalder, U. Collier, C. Dent, B. Epp, C. Hareesh Kumar *et al.*, "Renewables 2021-global status report," 2021.

[7] F. M. Graziani, "Directive of the European Parliament and of the Council on the promotion of the use of energy from renewable sources (recast)," *EASO: European Asylum Support Office*, 2017.

[8] G. T. Heydt, "The next generation of power distribution systems," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 225–235, 2010.

[9] SPD DSA Task Force. (2017) Dynamic Security Assessment (DSA), RG-CE System Protection & Dynamics Sub Group. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents/SOC%20documents/Regional_Groups_Continental_Europe/2017/DSA_REPORT_Public.pdf

[10] S. M. Zali and J. V. Milanović, "Generic model of active distribution network for large power system stability studies," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3126–3133, 2013.

[11] G. Chaspierre, G. Denis, P. Panciatici, and T. Van Cutsem, "An active distribution network equivalent derived from large-disturbance simulations with uncertainty," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4749–4759, 2020.

[12] G. Mitrentsis and H. Lens, "Data-driven dynamic models of active distribution networks using unsupervised learning techniques on field measurements," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 2952–2965, 2021.

[13] Y. Zhu and J. V. Milanović, "Automatic identification of power system load models based on field measurements," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3162–3171, 2017.

[14] P. Aristidou and T. Van Cutsem, "A parallel processing approach to dynamic simulations of combined transmission and distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 58–65, 2015.

[15] R. Venkatraman, S. K. Khaitan, and V. Ajjarapu, "Dynamic co-simulation methods for combined transmission-distribution system with integration time step impact on convergence," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1171–1181, 2018.

[16] A. K. Bharati and V. Ajjarapu, "Investigation of relevant distribution system representation with DG for voltage stability margin assessment," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2072–2081, 2019.

[17] J. V. Milanovic, K. Yamashita, S. M. Villanueva, S. Ž. Djokic, and L. M. Korunović, "International industry practice on power system load modeling," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3038–3046, 2012.

[18] K. Yamashita, S. Djokic, J. Matevosyan, F. Resende, L. Korunovic, Z. Dong, and J. Milanovic, "Modelling and aggregation of loads in flexible power networks–scope and status of the work of CIGRE WG C4. 605," *IFAC Proceedings Volumes*, vol. 45, no. 21, pp. 405–410, 2012.

[19] I. Dzafic, M. Glavic, and S. Tesnjak, "A component-based power system model-driven architecture," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 2109–2110, 2004.

[20] D. Kosterev, A. Meklin, J. Undrill, B. Lesieutre, W. Price, D. Chassin, R. Bravo, and S. Yang, "Load modeling in power system studies: WECC progress update," in *IEEE Power and Energy Society General Meeting*. IEEE, 2008, pp. 1–8.

[21] L. Zhu, X. Li, H. Ouyang, Y. Wang, W. Liu, and K. Shao, "Research on component-based approach load modeling based on energy management sys-

tem and load control system," in *IEEE PES Innovative Smart Grid Technologies*. IEEE, 2012, pp. 1–6.

[22] (2012) WECC MVWG Load Model Report Ver. 1.0. [Online]. Available: https://www.wecc.org/Reliability/WECC%20MVWG%20Load%20Model%20Report%20ver%201%200.pdf

[23] A. Gaikwad, P. Markham, and P. Pourbeik, "Implementation of the wecc composite load model for utilities using the component-based modeling approach," in *IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*. IEEE, 2016, pp. 1–5.

[24] H. Renmu, M. Jin, and D. J. Hill, "Composite load modeling via measurement approach," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 663–672, 2006.

[25] B.-K. Choi, H.-D. Chiang, Y. Li, H. Li, Y.-T. Chen, D.-H. Huang, and M. G. Lauby, "Measurement-based dynamic load models: Derivation, comparison, and validation," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1276–1283, 2006.

[26] M. Jin, H. Renmu, and D. J. Hill, "Load modeling by finding support vectors of load data from field measurements," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 726–735, 2006.

[27] Y. Liu, K. Sun, and Y. Liu, "A measurement-based power system model for dynamic response estimation and instability warning," *Electric Power Systems Research*, vol. 124, pp. 1–9, 2015.

[28] I. F. Visconti, D. A. Lima, J. M. C. de Sousa Costa, and N. R. de BC Sobrinho, "Measurement-based load modeling using transfer functions for dynamic simulations," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 111–120, 2013.

[29] E. O. Kontis, T. A. Papadopoulos, A. I. Chrysochos, and G. K. Papagiannis, "Measurement-based dynamic load modeling using the vector fitting technique," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 338–351, 2017.

[30] B. Zaker, G. B. Gharehpetian, and M. Karrari, "A novel measurement-based dynamic equivalent model of grid-connected microgrids," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2032–2043, 2018.

[31] K. S. Metallinos, T. A. Papadopoulos, and C. A. Charalambous, "Derivation and evaluation of generic measurement-based dynamic load models," *Electric Power Systems Research*, vol. 140, pp. 193–200, 2016.

[32] J. V. Milanović and S. M. Zali, "Validation of equivalent dynamic model of active distribution network cell," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2101–2110, 2012.

[33] F. Conte, F. D'Agostino, and F. Silvestro, "Operational constrained nonlinear modeling and identification of active distribution networks," *Electric Power Systems Research*, vol. 168, pp. 92–104, 2019.

[34] X. Shang, Z. Li, J. Zheng, and Q. Wu, "Equivalent modeling of active distribution network considering the spatial uncertainty of renewable energy resources," *International Journal of Electrical Power & Energy Systems*, vol. 112, pp. 83–91, 2019.

[35] C. Wang, Z. Wang, J. Wang, and D. Zhao, "Robust time-varying parameter identification for composite load modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 967–979, 2017.

[36] N. Fulgêncio, C. Moreira, L. Carvalho, and J. P. Lopes, "Aggregated dynamic model of active distribution networks for large voltage disturbances," *Electric Power Systems Research*, vol. 178, p. 106006, 2020.

[37] E. O. Kontis, T. A. Papadopoulos, M. H. Syed, E. Guillo-Sansano, G. M. Burt, and G. K. Papagiannis, "Artificial-intelligence method for the derivation of generic aggregated dynamic equivalent models," *IEEE Transactions on Power Systems*, 2019.

[38] D. Han, J. Ma, R.-M. He, and Z.-Y. Dong, "A real application of measurement-based load modeling in large-scale power grids and its validation," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1756–1764, 2009.

[39] D. P. Stojanović, L. M. Korunović, and J. Milanović, "Dynamic load modelling based on measurements in medium voltage distribution network," *Electric Power Systems Research*, vol. 78, no. 2, pp. 228–238, 2008.

[40] X. Tang, K. N. Hasan, J. V. Milanović, K. Bailey, and S. J. Stott, "Estimation and validation of characteristic load profile through smart grid trials in a medium voltage distribution network," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1848–1859, 2017.

[41] G. Mitrentsis and H. Lens, "A dynamic active distribution network equivalent for enhancing the generalization capability of the exponential recovery model in stability studies," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2709–2712, 2021.

[42] D. N. Kosterev, C. W. Taylor, and W. A. Mittelstadt, "Model validation for the August 10, 1996 WSCC system outage," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 967–979, 1999.

[43] S. A. Arefifar and W. Xu, "Online tracking of voltage-dependent load parameters using ULTC created disturbances," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 130–139, 2012.

[44] A. Bokhari, A. Alkan, R. Dogan, M. Diaz-Aguiló, F. De Leon, D. Czarkowski, Z. Zabar, L. Birenbaum, A. Noel, and R. E. Uosef, "Ex-

perimental determination of the ZIP coefficients for modern residential, commercial, and industrial loads," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1372–1381, 2013.

[45] L. M. Hajagos and B. Danai, "Laboratory measurements and models of modern loads and their effect on voltage stability studies," *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 584–592, 1998.

[46] N. Lu, Y. Xie, Z. Huang, F. Puyleart, and S. Yang, "Load component database of household appliances and small office equipment," in *IEEE Power and Energy Society General Meeting*. IEEE, 2008, pp. 1–5.

[47] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*. New York: McGraw-Hill, 1994, vol. 7.

[48] W. Price, H. Chiang, H. Clark, C. Concordia, D. Lee, J. Hsu, S. Ihara, C. King, C. Lin, Y. Mansour *et al.*, "Load representation for dynamic performance analysis," *IEEE Transactions on Power Systems*, vol. 8, no. 2, 1993.

[49] S.-H. Lee, S.-E. Son, S.-M. Lee, J.-M. Cho, K.-B. Song, and J.-W. Park, "Kalman-filter based static load modeling of real power system using k-ems data," *Journal of Electrical Engineering and Technology*, vol. 7, no. 3, pp. 304–311, 2012.

[50] Y. Ge, A. J. Flueck, D.-K. Kim, J.-B. Ahn, J.-D. Lee, and D.-Y. Kwon, "An event-oriented method for online load modeling based on synchrophasor data," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 2060–2068, 2015.

[51] A. S. Carneiro, L. F. Araujo, J. L. R. Pereira, P. A. Garcia, I. D. Melo, and M. B. Amaral, "Static load modeling based on field measurements," in *IEEE Manchester PowerTech*. IEEE, 2017, pp. 1–5.

[52] D. He, T. Habetler, M. J. Mousavi, and N. Kang, "A ZIP model-based feeder load modeling and forecasting method," in *IEEE Power and Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[53] M. W. Asres, A. A. Girmay, C. Camarda, and G. T. Tesfamariam, "Non-intrusive load composition estimation from aggregate ZIP load models using machine learning," *International Journal of Electrical Power & Energy Systems*, vol. 105, pp. 191–200, 2019.

[54] S. M. H. Rizvi, S. K. Sadanandan, and A. K. Srivastava, "Real-time ZIP load parameter tracking using sensitivity-based adaptive window and variable elimination with realistic synchrophasor data," *IEEE Transactions on Industry Applications*, vol. 57, no. 6, pp. 6525–6536, 2021.

[55] C. Wang, Z. Wang, and S. Ma, "SVM-based parameter identification for static load modeling," in *IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*. IEEE, 2018, pp. 1–5.

[56] S. Pandey, A. K. Srivastava, P. Markham, M. Patel *et al.*, "Online estimation of steady-state load models considering data anomalies," *IEEE Transactions on Industry Applications*, vol. 54, no. 1, pp. 712–721, 2017.

[57] P. Regulski, D. Vilchis-Rodriguez, S. Djurović, and V. Terzija, "Estimation of composite load model parameters using an improved particle swarm optimization method," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 553–560, 2014.

[58] L. M. Korunović, J. V. Milanović, S. Z. Djokic, K. Yamashita, S. M. Villanueva, and S. Sterpu, "Recommended parameter values and ranges of most frequently used static load models," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 5923–5934, 2018.

[59] M. Leinakse, P. Tani, and J. Kilter, "Impact of distributed generation on estimation of exponential load models," in *2019 IEEE Power and Energy Society General Meeting*. IEEE, 2019, pp. 1–5.

[60] B.-K. Choi and H.-D. Chiang, "Multiple solutions and plateau phenomenon in measurement-based load model development: Issues and suggestions," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 824–831, 2009.

[61] A. Samui and S. Samantaray, "An active islanding detection scheme for inverter-based DG with frequency dependent ZIP–exponential static load model," *International Journal of Electrical Power & Energy Systems*, vol. 78, pp. 41–50, 2016.

[62] M. Sadeghi *et al.*, "Determination of ZIP parameters with least squares optimization method," in *2009 IEEE Electrical Power & Energy Conference (EPEC)*. IEEE, 2009, pp. 1–6.

[63] W. W. Price, K. A. Wirgau, A. Murdoch, J. V. Mitsche, E. Vaahedi, and M. El-Kady, "Load modeling for power flow and transient stability computer studies," *IEEE Transactions on Power Systems*, vol. 3, no. 1, pp. 180–187, 1988.

[64] J. Anton, R. Mathews, D. Perkins, M. Herndon, P. Dworsky, and T. Mathews, "Analysis of performance accelerator running ETMSP. Final report," Electric Power Research Inst., Palo Alto, CA (United States), Tech. Rep., 1993.

[65] E. Welfonder, H. Weber, and B. Hall, "Investigations of the frequency and voltage dependence of load part systems using a digital self-acting measuring and identification system," *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 19–25, 1989.

[66] H. Bai, P. Zhang, and V. Ajjarapu, "A novel parameter identification approach via hybrid learning for aggregate load modeling," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1145–1154, 2009.

[67] D. Karlsson and D. J. Hill, "Modelling and identification of nonlinear dynamic loads in power systems," *IEEE Transactions on Power Systems*, vol. 9, no. 1, pp. 157–166, 1994.

[68] D. J. Hill, "Nonlinear dynamic load models with recovery for voltage stability studies," *IEEE Transactions on Power Systems*, vol. 8, no. 1, pp. 166–176, 1993.

[69] A. Rouhani and A. Abur, "Real-time dynamic parameter estimation for an exponential dynamic load model," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1530–1536, 2015.

[70] J. Ma, D. Han, R.-M. He, Z.-Y. Dong, and D. J. Hill, "Reducing identified parameters of measurement-based composite load model," *IEEE Transactions on Power Systems*, vol. 23, no. 1, pp. 76–83, 2008.

[71] J.-K. Kim, K. An, J. Ma, J. Shin, K.-B. Song, J.-D. Park, J.-W. Park, and K. Hur, "Fast and reliable estimation of composite load model parameters using analytical similarity of parameter sensitivity," *IEEE transactions on Power Systems*, vol. 31, no. 1, pp. 663–671, 2015.

[72] A. Keyhani, W. Lu, and G. T. Heydt, "Composite neural network load models for power system stability analysis," in *IEEE PES Power Systems Conference and Exposition*. IEEE, 2004, pp. 1159–1163.

[73] M. Bostanci, J. Koplowitz, and C. Taylor, "Identification of power system load dynamics using artificial neural networks," *IEEE Transactions on Power Systems*, vol. 12, no. 4, pp. 1468–1473, 1997.

[74] B.-Y. Ku, R. J. Thomas, C.-Y. Chiou, and C.-J. Lin, "Power system dynamic load modeling using artificial neural networks," *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1868–1874, 1994.

[75] D. Mandic and J. Chambers, *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. Wiley, 2001.

[76] X. Feng, Z. Lubosny, and J. Bialek, "Dynamic equivalencing of distribution network with high penetration of distributed generation," in *Proceedings of the 41st International Universities Power Engineering Conference*, vol. 2. IEEE, 2006, pp. 467–471.

[77] P. N. Papadopoulos, T. A. Papadopoulos, P. Crolla, A. J. Roscoe, G. K. Papagiannis, and G. M. Burt, "Black-box dynamic equivalent model for micro-

grids using measurement data," *IET Generation, Transmission & Distribution*, vol. 8, no. 5, pp. 851–861, 2014.

[78] A. M. Azmy and I. Erlich, "Identification of dynamic equivalents for distribution power networks using recurrent anns," in *IEEE PES Power Systems Conference and Exposition*. IEEE, 2004, pp. 348–353.

[79] X. Feng, Z. Lubosny, and J. Bialek, "Identification based dynamic equivalencing," in *IEEE Lausanne Power Tech*. IEEE, 2007, pp. 267–272.

[80] A. Samadi, L. Söder, E. Shayesteh, and R. Eriksson, "Static equivalent of distribution grids with high penetration of pv systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1763–1774, 2015.

[81] G. Chaspierre, P. Panciatici, and T. Van Cutsem, "Aggregated dynamic equivalent of a distribution system hosting inverter-based generators," in *Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–7.

[82] K. Zhang, H. Zhu, and S. Guo, "Dependency analysis and improved parameter estimation for dynamic composite load modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3287–3297, 2016.

[83] Q. Liu, Y. Chen, and D. Duan, "The load modeling and parameters identification for voltage stability analysis," in *Proceedings. International Conference on Power System Technology*, vol. 4. IEEE, 2002, pp. 2030–2033.

[84] L. T. M. Mota and A. A. Mota, "Load modeling at electric power distribution substations using dynamic load parameters estimation," *International Journal of Electrical Power & Energy Systems*, vol. 26, no. 10, pp. 805–811, 2004.

[85] Z. Wang and J. Wang, "Time-varying stochastic assessment of conservation voltage reduction based on load modeling," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2321–2328, 2014.

[86] P. Ju, E. Handschin, and D. Karlsson, "Nonlinear dynamic load modelling: model and parameter estimation," *IEEE Transactions on Power Systems*, vol. 11, no. 4, pp. 1689–1697, 1996.

[87] P. Jazayeri, W. Rosehart, and D. T. Westwick, "A multistage algorithm for identification of nonlinear aggregate power system loads," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1072–1079, 2007.

[88] J. De Kock, F. Van Der Merwe, and H. Vermeulen, "Induction motor parameter estimation through an output error technique," *IEEE Transactions on Energy Conversion*, vol. 9, no. 1, pp. 69–76, 1994.

[89] S. Kamoun and R. Malhamé, "Convergence characteristics of a maximum likelihood load model identification scheme," *Automatica*, vol. 28, no. 5, pp. 885–896, 1992.

[90] I. A. Hiskens, "Nonlinear dynamic model evaluation from disturbance measurements," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 702–710, 2001.

[91] M. Khodayar and J. Wang, "Probabilistic time-varying parameter identification for load modeling: A deep generative approach," *IEEE Transactions on Industrial Informatics*, 2020.

[92] J. Xie, Z. Ma, K. Dehghanpour, Z. Wang, Y. Wang, R. Diao, and D. Shi, "Imitation and transfer Q-learning-based parameter identification for composite load modeling," *IEEE Transactions on Smart Grid*, 2020.

[93] X. Wang, Y. Wang, D. Shi, J. Wang, and Z. Wang, "Two-stage WECC composite load modeling: A double deep Q-learning networks approach," *IEEE Transactions on Smart Grid*, 2020.

[94] M. Cui, M. Khodayar, C. Chen, X. Wang, Y. Zhang, and M. E. Khodayar, "Deep learning-based time-varying parameter identification for system-wide

load modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6102–6114, 2019.

[95] F. Resende, J. Matevosyan, and J. Milanovic, "Application of dynamic equivalence techniques to derive aggregated models of active distribution network cells and microgrids," in *IEEE Grenoble Conference*, 2013.

[96] F. C. Véliz, S. Gomes, S. L. Varricchio, and S. Carneiro, "Dynamic equivalents for large ac networks using modal analysis for s-domain models," in *IEEE Power Engineering Society General Meeting*. IEEE, 2005, pp. 132–138.

[97] D. Bonvin and D. Mellichamp, "A unified derivation and critical review of modal approaches to model reduction," *International Journal of Control*, vol. 35, no. 5, pp. 829–848, 1982.

[98] L. A. Aguirre, "Quantitative measure of modal dominance for continuous systems," in *Proceedings of 32nd IEEE Conference on Decision and Control*. IEEE, 1993, pp. 2405–2410.

[99] J. Undrill and A. Turner, "Construction of power system electromechanical equivalents by modal analysis," *IEEE Transactions on Power Apparatus and Systems*, no. 5, pp. 2049–2059, 1971.

[100] J. Undrill, J. Casazza, E. Gulachenski, and L. Kirchrnayer, "Electromechanical equivalent for use in power system stability studies," *IEEE Transactions on Power Apparatus and Systems*, no. 5, pp. 2060–2071, 1971.

[101] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Transactions on Automatic Control*, vol. 26, no. 1, pp. 17–32, 1981.

[102] K. Glover, "All optimal hankel-norm approximations of linear multivariable systems and their L∞-error bounds," *International journal of control*, vol. 39, no. 6, pp. 1115–1193, 1984.

[103] D. Chaniotis and M. Pai, "Model reduction in power systems using Krylov subspace methods," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 888–894, 2005.

[104] A. M. Zin, B. Kok, M. Mustafa, K. Lo, and A. Ariffin, "Time domain dynamic aggregation of generating unit based on structure preserving approach," in *Proceedings. National Power Engineering Conference*. IEEE, 2003, pp. 154–160.

[105] M. Ourari, L.-A. Dessaint, and V.-Q. Do, "Generating units aggregation for dynamic equivalent of large power systems," in *IEEE Power Engineering Society General Meeting.* IEEE, 2004, pp. 1535–1541.

[106] G. N. Ramaswamy, L. Rouco, O. Fillatre, G. C. Verghese, P. Panciatici, B. C. Lesieutre, and D. Peltier, "Synchronic modal equivalencing (sme) for structure-preserving dynamic equivalents," *IEEE Transactions on Power Systems*, vol. 11, no. 1, pp. 19–29, 1996.

[107] J. H. Chow, R. Galarza, P. Accari, and W. W. Price, "Inertial and slow coherency aggregation algorithms for power system dynamic model reduction," *IEEE Transactions on Power Systems*, vol. 10, no. 2, pp. 680–685, 1995.

[108] A. J. Germond and R. Podmore, "Dynamic aggregation of generating unit models," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 4, pp. 1060–1069, 1978.

[109] M. G. Ippolito, R. Musca, and G. Zizzo, "Analysis and simulations of the primary frequency control during a system split in continental Europe power system," *Energies*, vol. 14, no. 5, p. 1456, 2021.

[110] User Manual: Power Quality Network Analyser Model PQI-DA smart. [Online]. Available: https://www.a-eberle.de/wp-content/uploads/2021/02/BA_PQI-DA-smart_EN.pdf

[111] Operating Manual: Network Analyzer PQ-Box 150. [Online]. Available: https://www.a-eberle.de/wp-content/uploads/2021/02/IM_PQBox150_EN_Hardware_584.0841.pdf

[112] DWD Climate Data Center (CDC):. [Online]. Available: https://opendata.dwd.de/climate_environment/CDC/

[113] R. Ahmed, V. Sreeram, Y. Mishra, and M. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109792, 2020.

[114] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Conversion and Management*, vol. 195, pp. 328–345, 2019.

[115] G. Mitrentsis and H. Lens, "Unsupervised learning method for clustering dynamic behavior in the context of power systems," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 13 024–13 029, 2020.

[116] ——, "Dynamic modeling of active distribution networks using cluster analysis of field measurement data," in *NEIS 2020; Conference on Sustainable Energy Supply and Energy Storage Systems*. VDE, 2020, pp. 1–7.

[117] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[118] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[119] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[120] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[121] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *International Conference on Knowledge Discovery and Data Mining*, vol. 96, no. 34, Portland, OR, USA, 1996, pp. 226–231.

[122] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

[123] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, "A review of classification problems and algorithms in renewable energy applications," *Energies*, vol. 9, no. 8, p. 607, 2016.

[124] K. Mets, F. Depuydt, and C. Develder, "Two-stage load pattern clustering using fast wavelet transformation," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2250–2259, 2015.

[125] R. C. De Amorim and C. Hennig, "Recovering the number of clusters in data sets with noise features using feature rescaling factors," *Information Sciences*, vol. 324, pp. 126–145, 2015.

[126] C. Chen, J. Wang, Z. Li, H. Sun, and Z. Wang, "PMU uncertainty quantification in voltage stability analysis," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 2196–2197, 2014.

[127] K. Yamashita, M. Asada, and K. Yoshimura, "A development of dynamic load model parameter derivation method," in *IEEE Power and Energy Society General Meeting*. Calgary, Canada: IEEE, 2009, pp. 1–8.

[128] L. Ljung, *System identification: Theory for the user*. PTR Prentice Hall, Upper Saddle River, NJ, 1999.

[129] Commission Regulation (EU) 2016/631 of 14 April 2016 establishing a network code on requirements for grid connection of generators. [Online]. Available: http://data.europa.eu/eli/reg/2016/631/oj

[130] B. Hall, "Experimentelle Untersuchung zur frequenz- und spannungsabhängigen Lastaufnahme elektrischer Verbraucherteilnetze," Ph.D. dissertation, University of Stuttgart, 1993.

[131] H. Garnier, M. Mensler, and A. Richard, "Continuous-time model identification from sampled data: implementation issues and performance evaluation," *International Journal of Control*, vol. 76, no. 13, pp. 1337–1357, 2003.

[132] V. Knyazkin, C. A. Canizares, and L. H. Soder, "On the parameter estimation and modeling of aggregate power system loads," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1023–1031, 2004.

[133] D. E. Goldberg, "Genetic algorithms in search," *Optimization, and Machine-Learning*, 1989.

[134] H. Garnier, L. Wang, and P. C. Young, "Direct identification of continuous-time models from sampled data: Issues, basic solutions and relevance," in *Identification of continuous-time models from sampled data*. Springer, 2008, pp. 1–29.

[135] D. P. Kingma and J. Ba. (2014) Adam: A method for stochastic optimization. [Online]. Available: https://arxiv.org/abs/1412.6980

[136] G. Mitrentsis and H. Lens, "Probabilistic dynamic model of active distribution networks using gaussian processes," in *IEEE Madrid PowerTech*. IEEE, 2021, pp. 1–6.

[137] ——, "A Gaussian process framework for the probabilistic dynamic modeling of active distribution networks using exogenous variables," *Electric Power Systems Research*, vol. 211, p. 108403, 2022.

[138] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[139] J. Kocijan, *Modelling and control of dynamic systems using Gaussian process models*. Springer, 2016.

[140] J. Kocijan, A. Girard, B. Banko, and R. Murray-Smith, "Dynamic systems identification with Gaussian processes," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 11, no. 4, pp. 411–424, 2005.

[141] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and Gaussian processes—revisited," *Automatica*, vol. 48, no. 8, pp. 1525–1535, 2012.

[142] A. Svensson and T. B. Schön, "A flexible state–space model for learning nonlinear dynamical systems," *Automatica*, vol. 80, pp. 189–199, 2017.

[143] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[144] W. Price, C. Taylor, and G. Rogers, "Standard load models for power flow and dynamic performance simulation," *IEEE Transactions on Power Systems*, vol. 10, no. CONF-940702-, 1995.

[145] L. Pereira, D. Kosterev, P. Mackin, D. Davies, J. Undrill, and W. Zhu, "An interim dynamic induction motor model for stability studies in the WSCC," *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 1108–1115, 2002.

[146] A. Ellis, D. Kosterev, and A. Meklin, "Dynamic load models: Where are we?" in *Transmission and Distribution Conference and Exhibition*, no. 2. IEEE Brightun, England, 2006, pp. 1320–1324.

[147] U. D. Annakkage, N.-K. C. Nair, Y. Liang, A. Gole, V. Dinavahi, B. Gustavsen, T. Noda, H. Ghasemi, A. Monti, M. Matar *et al.*, "Dynamic system equivalents: A survey of available techniques," *IEEE Transactions on Power Delivery*, vol. 27, no. 1, pp. 411–420, 2011.

[148] DIgSILENT Powerfactory, "Powerfactory user's manual," *DIgSILENT, GmbH*, 2021.

[149] NEPLAN, "Neplan user's guide," *NEPLAN AG*, 2021.

[150] Simulink Documentation, "Simulation and model-based design," *MathWorks*, 2021.

[151] G. Mitrentsis and H. Lens, "An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting," *Applied Energy*, vol. 309, p. 118473, 2022.

[152] G. Mitrentsis, M. Liu, and H. Lens, "Open source tool for probabilistic short-term PV and wind power forecasting," in *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, 2022, pp. 1–6.

[153] A. Ajagekar and F. You, "Quantum computing for energy systems optimization: Challenges and opportunities," *Energy*, vol. 179, pp. 76–89, 2019.

[154] X. Pan, T. Zhao, M. Chen, and S. Zhang, "Deepopf: A deep neural network approach for security-constrained DC optimal power flow," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 1725–1735, 2020.

[155] E. Yeung, S. Kundu, and N. Hodas, "Learning deep neural network representations for koopman operators of nonlinear dynamical systems," in *American Control Conference (ACC)*. IEEE, 2019, pp. 4832–4839.

[156] K. Lee and K. T. Carlberg, "Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders," *Journal of Computational Physics*, vol. 404, p. 108973, 2020.

[157] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," *Nature communications*, vol. 9, no. 1, pp. 1–10, 2018.

[158] W. Sun, M. Zamani, M. R. Hesamzadeh, and H.-T. Zhang, "Data-driven probabilistic optimal power flow with nonparametric Bayesian modeling and inference," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1077–1090, 2019.

[159] W. Lin, Z. Yang, J. Yu, S. Bao, and W. Dai, "Toward fast calculation of probabilistic optimal power flow," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3286–3288, 2019.

[160] W. Sun, M. Zamani, H.-T. Zhang, and Y. Li, "Probabilistic optimal power flow with correlated wind power uncertainty via Markov chain quasi-Monte-Carlo sampling," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 11, pp. 6058–6069, 2019.

[161] A. Venzke and S. Chatzivasileiadis, "Convex relaxations of probabilistic AC optimal power flow for interconnected AC and HVDC grids," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2706–2718, 2019.

[162] M. Adeen and F. Milano, "Modeling of correlated stochastic processes for the transient stability analysis of power systems," *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 4445–4456, 2021.

[163] ——, "Stochastic aggregated dynamic model of wind generation with correlated wind speeds," *Electric Power Systems Research*, vol. 212, p. 108312, 2022.

[164] I. D. Pasiopoulou, E. O. Kontis, T. Papadopoulos, and G. Papagiannis, "Effect of load modeling on power system stability studies," *Electric Power Systems Research*, vol. 207, p. 107846, 2022.