Anomaly Detection and Synthetic Data Generation for Power Systems Using Autoencoder Neural Networks

Wang, C.

DOI
10.4233/uuid:12708aca-dff2-4d59-aa0e-af5c275aa728

Publication date
2023

Document Version
Final published version

Citation (APA)
Wang, C. (2023). *Anomaly Detection and Synthetic Data Generation for Power Systems Using Autoencoder Neural Networks*. [Dissertation (TU Delft), Delft University of Technology].
https://doi.org/10.4233/uuid:12708aca-dff2-4d59-aa0e-af5c275aa728

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Anomaly Detection and Synthetic Data Generation for Power Systems Using Autoencoder Neural Networks

ANOMALY DETECTION AND SYNTHETIC DATA GENERATION FOR POWER SYSTEMS USING AUTOENCODER NEURAL NETWORKS

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Wednesday 22 March 2023 at 17:30 o'clock

by

Chenguang WANG

Master of Science in Electrical Engineering
Xi'an Jiaotong University, Xi'an, China
born in Jingzhou, Hubei Province, China

This dissertation has been approved by the promotors.

Promotor: Prof.dr. P. Palensky
Copromotor: Dr. S.H. Tindemans

Composition of the doctoral committee:

Rector Magnificus,	Chairperson
Prof.dr. P. Palensky	Delft University of Technology, promotor
Dr. S.H. Tindemans	Delft University of Technology, copromotor

Independent members:

Prof.dr.ir. J.A. La Poutré	Delft University of Technology
Dr. P.H. Nguyen	Eindhoven University of Technology
Dr. P. Mohajerin Esfahani	Delft University of Technology
Dr. S. Chatzivasileiadis	Technical University of Denmark
Dr.ir. S.E. Verwer	Delft University of Technology
Prof.dr.ir. P. Bauer	Delft University of Technology, reserve member

This research was financially supported by the China Scholarship Council.

Keywords: Anomaly Detection, Synthetic Data Generation, Autoencoder, Power System Operation and Planning, Machine Learning

Cover design by: Ting Hu

Printed by: Proefschriftenprinten.nl

ISBN 978-90-833109-4-7

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Dedicated to my family and my love

Chenguang Wang

NOTATION

List of Abbreviations

ACC	Accuracy
Adam	Adaptive moment estimation
AE	Autoencoder
am	Anomaly magnitude
AUC	Area under the curve
BDD	Bad data detection
CET	Central European time
(C)GAN	(Conditional) Generative adversarial network
CNN	Convolutional neural network
(C)VAE	(Conditional) Variational autoencoder
EENS	Expected energy not served
EM	Expectation-maximization
EMS	Energy management system
FN	False negative
FP	False positive
GMM	Gaussian mixture model
ICS	Industrial control system
ICT	Information and communication technology
IS	Importance sampling
K-NN	K-nearest neighbor
K-S	Kolmogorov-Smirnov
LOF	Local outlier factor
LOLE	Loss of load expectation
MC	Monte Carlo
MILP	Mixed integer linear program

OCR	Ordinal consistency rate
OVAE	Oriented variational autoencoder
PCA	Principal component analysis
PINN	Physics-informed neural network
PLC	Programmable logic controller
PPV	Precision
RMS	Root mean square
RNN	Recurrent neural network
ROC	Receiver operating characteristic curve
SCADA	Supervisory control and data acquisition
SVM	Support vector machine
TN	True negative
TP	True positive
UTC	Coordinated universal time
WLS	Weighted least squares

CONTENTS

Summary xiii

Samenvatting xvii

1 Introduction 1

1.1 Research Background and Focus . . . 2
1.1.1 Data-driven Power System Operations . . . 2
1.1.2 Research Focus: Anomaly Detection and Data Generation . . . 2
1.2 Challenges and Research Questions . . . 6
1.3 Contributions and Thesis Outline . . . 10

2 Autoencoder-based Anomaly Detection 15

2.1 Introduction . . . 17
2.1.1 State of the Art of Anomaly Detection . . . 17
2.1.2 Contribution and Outline . . . 19
2.2 Anomalous Data Attack Scenarios . . . 20
2.2.1 Power System State Estimation . . . 20
2.2.2 Stealth False Data Injection Attacks . . . 21
2.3 Autoencoder-based Anomaly Detectors . . . 22
2.3.1 Detector Schematic . . . 22
2.3.2 Anomaly Detection Mechanism . . . 24
2.4 Data Attack Detection Using Autoencoder . . . 25
2.4.1 Test System Modeling . . . 25
2.4.2 Detection Performance Analysis . . . 28
2.5 Detector Training Strategy . . . 31
2.5.1 Hyperparameter Tuning . . . 32
2.5.2 Threshold Selection Strategy Investigation . . . 34
2.5.3 Threshold Selection Strategy . . . 35

3 Anomaly Detector Performance Improvement **37**
3.1 Introduction . . . 38
3.1.1 Related Work and Motivation . . . 38
3.1.2 Contribution and Outline . . . 39
3.2 Detector Enhancements . . . 40
3.2.1 Problem Formulation . . . 40
3.2.2 Data Whitening Schemes . . . 41
3.2.3 Anomaly Localization Metrics . . . 44
3.2.4 Design of the Anomaly Detector . . . 45
3.3 Detection of Anomalous Wind Farm Generations . . . 47
3.3.1 Experiment Scenario Formulation . . . 47
3.3.2 Impact of Whitening Transformation . . . 48
3.3.3 Anomaly Detection Performance Evaluation . . . 49
3.3.4 Anomaly Localization Performance Evaluation . . . 52

4 Data Generation Using a Conditional Variational Autoencoder **57**
4.1 Introduction . . . 59
4.1.1 State of the Art of Data Generators . . . 59
4.1.2 Contribution and Outline . . . 60
4.2 Data Generation Mechanism . . . 61
4.2.1 CVAE-based Generative Model . . . 61
4.2.2 Training and Generation Process . . . 63
4.2.3 Generator Optimization Strategy . . . 64
4.3 Case Study on Country Level Load Data . . . 65
4.3.1 Data Source and Generation . . . 65
4.3.2 Data Quality Metrics . . . 66
4.3.3 Experimental Results Analysis . . . 68
4.4 Case Study on Multi-area Adequacy Assessment . . . 72
4.4.1 Multi-area Adequacy Assessment Structure . . . 73
4.4.2 Power System Model . . . 74
4.4.3 Multi-area Adequacy Assessment Results . . . 74
4.5 Case Study on Load Data of Individual Customers . . . 75
4.5.1 Data Source and Generation . . . 77
4.5.2 Experimental Results Analysis . . . 80

5 Controllable Generator: An Oriented Variational Autoencoder 87
5.1 Introduction . . . 89
5.1.1 Motivation and Related Work . . . 89
5.1.2 Contribution and Outline . . . 90
5.2 Data Generation Mechanism . . . 91
5.2.1 Importance Sampling for Risk Assessment . . . 91
5.2.2 Proposed Oriented VAE-Based Generative Model . . . 92
5.2.3 Importance Sampling with OVAE . . . 95
5.3 Case Study Description . . . 96
5.3.1 Electricity Demand Data and OVAE Model Structure . . . 96
5.3.2 Resource Adequacy Model . . . 97
5.3.3 Multi-Area Resource Adequacy Impacts . . . 98
5.4 Experimental Results of OVAE . . . 99
5.4.1 Impact of Extra Oriented Loss $\mathscr{L}_{Ori}$ on Model Training . . . 99
5.4.2 Validation of Latent Space Alignment . . . 100
5.4.3 Unbiased and Biased Sampling . . . 101
5.4.4 Quality Evaluation of Generated Data . . . 102
5.4.5 Effectiveness of Performing Semi-Supervised Learning . . . 104
5.4.6 Multi-Area Adequacy Assessment Results . . . 104

6 Conclusion and Recommendations 109
6.1 Conclusions . . . 110
6.2 Discussion and Research Recommendations . . . 113

Bibliography 117

Curriculum Vitæ 129

List of Publications 131

Acknowledgements 133

SUMMARY

The scale of the power system has been significantly expanded in recent decades. To gain real-time insights into the power system, an increasing number of sensors have been deployed to monitor grid states, resulting in a rapidly growing number of measurement points. Simultaneously, there has also been a rise in the penetration of renewable energy generation, with energy production that is highly variable and exhibits strong interdependence between different production locations. Such interdependence also applies to electricity demand at various network positions. Furthermore, new demand-side response strategies and policies enhance the flexibility of the power system, leading to changes in load profiles. These developments, combined with the structure of the network itself, mean that measurements in the power system generally exhibit strong dependencies. This dependency means that if you know one or more values, you can infer information about others. This applies to time series with measurements that follow each other chronologically as well as to snapshots that show different states of the system at a particular moment in time. A large collection of such time series and snapshots can be represented as a probability distribution in a multidimensional data space. While larger numbers of measurements enable smarter grid operations, high-dimensional stochastic variables with complex univariate and multivariate distributions could also complicate tasks in modeling power system data.

For some power system tasks, it is critical to model the distribution of measurements. Two examples are power system anomaly detection and synthetic data generation. Anomaly detection is vital for power system stability and economic dispatching. This thesis focuses on detecting anomalous measurements that physically make sense but represent uncommon states. Specifically, we aim to detect intentional anomalies in the form of stealthy power system data integrity attacks. Efforts have also been made to detect more general unintentional anomalies in renewable energy systems, such as mild reductions in the power output of wind turbines. The main challenge in detecting anomalies is that their rarity in historical data makes it difficult to explicitly model the pattern of anomalies. And even if anomaly patterns are well modeled, the detector only

learns to detect known anomalies, which is a significant weakness for novel anomalies. A second challenge is that the probability distributions of energy demand and renewable generation cannot easily be captured in a mathematical equation. This also poses a problem for the planning of power systems and calibration of operational tools, where generating synthetic measurement data is essential to analyze system performance in a large range of representative scenarios, especially when the available historical data is limited. To explore potential solutions, this thesis investigates the generation of electric energy demand at the national level and load profiles of individual customers. The challenge in generating data lies in reproducing both marginal distributions and multivariate dependencies from historical data. In addition, it is useful if a data generator is able to generate data with user-defined characteristics. Moreover, with existing data generators, validation of the quality of generated data is often limited to visual comparisons, which is not intuitive for power system data in a non-pictorial form.

Motivated by these challenges, this thesis contributes to power system anomaly detection and synthetic data generation by proposing an enhanced anomaly detector and novel data generator based on an autoencoder neural network and the related variational autoencoder.

First, for the detection of power system data attacks in high dimensional space with a highly unbalanced historical data set, an autoencoder neural network-based anomaly detector is proposed to reduce the dimension of power system measurements and learn data patterns. The detector considers anomaly detection as a one-class classification task by acquiring the dependencies intrinsic in 'normal' operation data only. Rare anomalies which deviate from patterns learned from normal states are then detected. By focusing on 'normal' operating conditions, it effectively overcomes the challenge of unbalanced training data that is inherent in power system attack detection and could be prepared for novel data attacks conducted by resourceful attackers. The performance of the proposed detector is validated using case studies based on the IEEE 118-bus system. The experiments demonstrate that the mechanism is able to robustly detect stealthy anomalies under a variety of attack scenarios. To guide other researchers to implement an autoencoder-based anomaly detector, the influence of the hyperparameter selection on the training and anomaly detection performance is investigated. The experimental results demonstrate that under proper configurations, the mechanism can demonstrate satisfactory learning efficiency and attack detection performance. Based on those results, preliminary hyperparameter selection and tuning strategies are put forward.

Second, the autoencoder neural network-based anomaly detector is improved to detect more general anomalies in renewable energy scenarios. Specifically, autoencoder neural networks are a powerful tool for the detection of unknown anomalies. A threshold for the (Euclidean) length of the residuals is typically used to identify anomalous states of a system. Correlation between residuals is identified as a source of misclassification. Thus, to accurately detect and localize the source of anomalies, whitening transformations that decorrelate input features and/or residuals are implemented. For a use case of distributed wind power generation, the performance of various data processing combinations is quantified. Whitening of the input data is found to be most beneficial for accurate detection, with a slight benefit for the combined whitening of inputs and residuals. For localization of anomalies, three anomaly localization metrics are proposed to quantify the dependability of the anomaly detector, measuring the degree of standout of anomalous dimensions and their difference from normal dimensions. It is found that the whitening of residuals is preferred for anomaly localization, and the best performance is obtained using standardization of the input data and whitening of the residuals using the *ZCA* or *ZCA-cor* whitening matrix with a small additional offset.

Third, this thesis moves one step further to investigate the performance of conditional variational autoencoder (CVAE)- and variational autoencoder (VAE)-based models (variants of the regular autoencoder) to generate multivariate load states. Going beyond common (C)VAE implementations, the model includes a stochastic variation of output samples under given latent vectors and co-optimizes the parameters for this output variability. The generation performance is evaluated using univariate and multivariate performance metrics. It is shown that the inclusion of output variation improves the statistical properties of the generated data. A Monte Carlo generation adequacy study on the European network is implemented to illustrate the models' ability to generate a realistic tail distribution of country-level load states. The experiments demonstrate that the proposed generator outperforms other data generation mechanisms on at least one statistical test and is competitive on all others. In addition to generating snapshots of country-level load states with limited diversity and variability, the models' capacity is also validated by generating synthetic load profiles representing a large variety of individual users, where the loads are at a lower aggregation level and therefore more stochastic. The experimental results demonstrate the CVAE model can capture temporal features of historical load profiles and generate 'realistic' data with satisfying univariate distributions and multivariate dependencies.

In the end, a novel oriented variational autoencoder (OVAE)-based generative model is proposed. Concretely, for many use cases, not just a data generator is needed, but also an ability to steer the type of samples that are generated. To do so, a connection could be established between the latent space code (where the data is sampled) and the generated data. Nevertheless, for a basic VAE model, the relation between the compressed latent space code and generated data is unconstrained, and thus it is difficult to infer properties of generated data by using specific sampled latent space codes. The unbiased generation of data could result in many redundant samples being generated and potentially in a time-consuming and non-trivial follow-up task to label and sieve generated data for particular applications. Given this, the OVAE model is proposed to establish a connection between latent space codes and generated data. The effectiveness of the OVAE model is evaluated at both the training and generation stages in a visual and statistical manner. Experimental results demonstrate that the efficiency of generating targeted samples is significantly improved by using an OVAE-based generator. Moreover, in order to lay the groundwork for dealing with situations where the analysis of certain data labels is time-consuming, e.g., risk labels of load states analyzed by Monte Carlo simulations, a semi-supervised learning scheme is proposed to train the OVAE model with incomplete labels. It is shown that even trained with incomplete data labels, the generative model is able to generate data with properties of interest that are correlated to latent space codes.

SAMENVATTING

De schaal van het energiesysteem is de laatste decennia aanzienlijk uitgebreid. Om real-time inzicht te krijgen in het energiesysteem zijn er steeds meer sensoren ingezet om de toestand van het net te bewaken, waardoor het aantal meetpunten snel groeit. Tegelijkertijd neemt de penetratie van hernieuwbare energiebronnen toe, met een energieproductie die zeer variabel is en een sterke afhankelijkheid vertoont tussen verschillende productielocaties. Een dergelijke afhankelijkheid geldt ook voor de elektriciteitsvraag op verschillende netwerklocaties. Tenslotte vergroten nieuwe vraagsturingsmechanismen de flexibiliteit van het elektriciteitssysteem, wat direct leidt tot veranderingen in belastingprofielen.

Deze ontwikkelingen, tezamen met de structuur van het netwerk zelf, beteken dat metingen in het energiesysteem over het algemeen sterke afhankelijkheden tussen meetwaardes vertonen. Deze afhankelijkheid betekent dat als je één of meerdere waardes kent, je informatie over andere waardes kunt afleiden. Dit geldt voor tijdreeksen met meetwaarden die chronologisch op elkaar volgen of voor momentopnamen die verschillende aspecten van het systeem op een bepaald moment in de tijd weergeven. Collectief kan een grote verzameling van dergelijke tijdreeksen en momentopnamen worden gerepresenteerd als een kansverdeling in een multidimensionale gegevensruimte. Hoewel grotere aantallen metingen slimmere netwerkoperaties mogelijk maken, worden de verdelingen van de bijbehorende kansverdelingen ook steeds complexer.

Voor sommige energiesysteemtaken is het van cruciaal belang de verdeling van de metingen te modelleren. Twee voorbeelden zijn het opsporen van anomalieën en het genereren van synthetische gegevens. Anomaliedetectie is van vitaal belang voor de stabiliteit van het energiesysteem en het efficiënt gebruik ervan. Dit proefschrift richt zich op het detecteren van afwijkende metingen die op technische grond weliswaar acceptabel zijn, maar (zeer) ongewone toestanden vertegenwoordigen. In het bijzonder willen we opzettelijke anomalieën detecteren in de vorm van verborgen aanvallen op de gegevensintegriteit van het energiesysteem. Daarnaast zijn de methodes ook toegepast om onopzettelijke anomalieën in hernieuwbare energiesystemen te detecteren, zoals een

lichte vermindering van de stroomproductie van windturbines. De belangrijkste uitdaging bij het opsporen van anomalieën is dat de zeldzaamheid ervan in historische gegevens het moeilijk maakt het patroon van anomalieën expliciet te modelleren. En zelfs als anomaliepatronen goed gemodelleerd zijn, leert de detector alleen bekende anomalieën te detecteren, wat een belangrijke zwakte is voor afwijkende anomalieën.

Een tweede uitdaging is dat de kansverdelingen van de energievraag en hernieuwbare opwek niet eenvoudig in een wiskundige vergelijking te vatten zijn. Dit is bijvoorbeeld problematisch voor de planning van energiesystemen en de kalibratie van operationele software, waarbij het genereren van synthetische meetgegevens essentieel is om de systeemprestaties in een groot aantal representatieve scenario's te analyseren, vooral wanneer historische gegevens niet toereikend zijn. In dit proefschrift wordt het genereren van de elektrische energievraag op landniveau en van belastingprofielen van individuele klanten onderzocht. De uitdaging bij het genereren van gegevens is dat een generator zowel marginale verdelingen als multivariate afhankelijkheden van historische gegevens moet reproduceren. Bovendien is het nuttig als een gegevensgenerator gegevens kan genereren met door de gebruiker gedefinieerde kenmerken. Bij bestaande data-generatoren blijft de validatie van de kwaliteit van gegenereerde gegevens vaak beperkt tot visuele vergelijkingen, wat minder goed toepasbaar is op gegevens uit elektriciteitsnetwerken.

Gemotiveerd door deze uitdagingen draagt dit proefschrift bij aan de opsporing van anomalieën in elektriciteitssystemen en het genereren van synthetische meetgegevens, daarbij gebruik makend van een *autoencoder* neuraal netwerk en de verwante *variationele autoencoder*.

De anomaliedetector op basis van een autoencoder neuraal netwerk is in staat om de dimensie van energiesysteemmetingen te reduceren en datapatronen te leren. Deze wordt eerst gebruikt voor de detectie van aanvallen op energiesysteemgegevens met een zeer onevenwichtige historische dataset. De detector beschouwt anomaliedetectie als een classificatietaak van één klasse door alleen de afhankelijkheden te leren die inherent zijn aan 'normale' gegevens. Toestanden die afwijken van patronen die behoren bij normale toestanden worden dan geclassificeerd als anomalie. Hierdoor wordt de uitdaging van onevenwichtige trainingsgegevens, die inherent is aan de detectie van aanvallen op het energiesysteem, effectief overwonnen en is het systeem voorbereid op onbekende aanvallen door vindingrijke aanvallers. De prestaties van de voorgestelde detector worden gevalideerd aan de hand van het IEEE 118-bussysteem. De experimenten tonen

aan dat het mechanisme in staat is om op robuuste wijze verborgen aanvallen te detecteren onder verscheidene aanvalsscenario's. Om andere onderzoekers te helpen bij het implementeren van deze anomaliedetector, wordt ook de invloed van de hyperparameterselectie op het trainingsproces en de detectieprestaties onderzocht. Op basis van deze resultaten worden voorlopige hyperparameterselectie- en afstemstrategieën voorgesteld.

Ten tweede wordt de anomaliedetector verbeterd om meer algemene anomalieën in duurzame energiescenario's te detecteren. Gewoonlijk wordt een drempelwaarde voor de (Euclidische) lengte van de residuen gebruikt om een toestand als al dan niet afwijkend te classificeren. We laten zien dat correlatie tussen elementen van residuen een belangrijke bron van misclassificatie is. Om anomalieën nauwkeuriger te detecteren en de bron ervan te lokaliseren, worden witmakende (*whitening*) transformaties toegepast die inputkenmerken en/of residuen decorreleren. Voor een specifiek voorbeen van gedistribueerde windenergieopwekking worden de prestaties van verschillende detectorconfiguraties gekwantificeerd. *Whitening* van de inputgegevens blijkt het gunstigst voor nauwkeurige detectie, met een gering voordeel voor de gecombineerde *whitening* van input en residuen. Voor de lokalisatie van anomalieën worden drie indicatoren voor anomalielokalisatie voorgesteld om de bruikbaarheid van de anomaliedetector voor dit doel te kwantificeren. De beste prestaties worden verkregen door standaardisatie van de invoergegevens en *whitening* van de residuen met behulp van de *ZCA* of *ZCA-cor* methodes met een kleine extra compensatie.

Ten derde gaat dit proefschrift een stap verder en onderzoekt de prestaties van modellen op basis van de conditionele variationele autoencoder (CVAE) en de variationele autoencoder (VAE) (varianten van de gewone autoencoder) om multivariate momentopnames van de elektriciteitsvraag te genereren. In toevoeging op gebruikelijke (C)VAE-implementaties injecteert dit model ruis in de gegenereerde data, op basis van medegeoptimaliseerde parameters. De prestaties worden geëvalueerd met indicatoren voor de kwaliteit van univariate en multivariate gelijkenis met de trainingsdata. Aangetoond wordt dat het opnemen van ruis de statistische eigenschappen van de gegenereerde data verbetert. Een Monte Carlo-studie naar de leveringszekerheid van het Europese netwerk wordt uitgevoerd om te illustreren dat de modellen in staat zijn om ook de staart van kansverdelingen accuraat te modelleren. De experimenten tonen aan dat de voorgestelde datagenerator op ten minste één statistische test beter presteert dan andere mechanismen voor het genereren van gegevens en op alle andere tests concurrerend

is. Naast het genereren van momentopnames van vraagtoestanden op landniveau met relatief beperkte diversiteit en variabiliteit, wordt de brede inzetbaarheid van de methode ook gevalideerd door het genereren van synthetische belastingprofielen die een grote verscheidenheid aan individuele gebruikers vertegenwoordigen. Deze meetgegevens hebben een lager aggregatieniveau en hebben daarom een inherent grotere variatie. De experimentele resultaten tonen aan dat het CVAE-model temporele kenmerken van historische belastingprofielen kan vastleggen en 'realistische' gegevens kan genereren met bevredigende univariate verdelingen en multivariate afhankelijkheden.

Tot slot wordt een nieuw generatief model geïntroduceerd, in de vorm van de georiënteerde variationele autoencoder (OVAE). Aanleiding hiervoor is dat voor veel toepassingen niet alleen een gegevensgenerator nodig is, maar ook de mogelijkheid om sturing te geven aan de kenmerken van gegenereerde metingen. Dat is in principe al mogelijk bij een standaard VAE-model, maar daarvoor is het nodig om een verband te leren tussen de latente code (waaruit de synthetische metingen worden gegenereerd) en de eigenschappen van de synthetische data – en dit verband kan zeer complex zijn. Een andere mogelijkheid is het genereren van aspecifieke synthetische data en het filteren daarvan op basis van de gewenste kenmerken. Het nadeel hiervan is dat dit een tijdrovende en niet-triviale taak kan zijn. Het OVAE-model maakt het mogelijk om al tijdens de training een verband te leggen tussen latente ruimtecodes en gegenereerde gegevens en zodoende op eenvoudige wijze de gewenste gegevens te genereren. De doeltreffendheid van het OVAE-model wordt zowel in de trainingsfase als in de generatiefase op visuele en statistische wijze geëvalueerd. Experimentele resultaten tonen aan dat de efficiëntie van het genereren van gerichte monsters aanzienlijk wordt verbeterd door het gebruik van een op OVAE gebaseerde generator. Om dit model ook effectief in te zetten voor situaties waarin de berekening van de kenmerkende labels tijdrovend is, zoals voor risicolabels op basis van Monte Carlo-simulaties, wordt bovendien een semi-gesuperviseerd leerschema voorgesteld om het OVAE-model te trainen met onvolledige labels. Aangetoond wordt dat het generatieve model, zelfs getraind met onvolledige gegevenslabels, in staat is om gegevens te genereren met specifieke eigenschappen die gecorreleerd zijn met latente codes.

1

INTRODUCTION

This chapter introduces the subjects of power system anomaly detection and data generation. Research motivation and background information are first stated, followed by two research focuses. For each research focus, research challenges and associated research questions are described and proposed, respectively. In the end, the contributions of this research are summarized and the structure of this thesis is given.

1

1.1. RESEARCH BACKGROUND AND FOCUS

1.1.1. DATA-DRIVEN POWER SYSTEM OPERATIONS

The scale of the power system has been significantly expanded during the last decades. For example, global electricity consumption increased from 18,695 TWh to 23,774 TWh between 2010 to 2019 [1]. Meanwhile, the installed generation capacity grew by 44.49% from 5.08 TW to 7.34 TW [2]. To have real-time insight into the power system, more and more meters have been deployed to monitor grid states, making measurements more numerous. It is reported that more than 653.30 million smart electricity meters were deployed in Asia at the end of 2019 [3].

The penetration of renewable energy generation is increasing. According to [2], from 2010 to 2019, the renewable electricity capacity increased by 109.77%, reaching 2.49 TW. This is equivalent to more than half of the global generation capacity of nuclear power (0.37 TW) and fossil fuels (4.38 TW). Solar and wind are two primary renewable energy resources. However, these resources are variable, making the power generated by solar panels and wind turbines highly unpredictable. In addition, wind speed and solar irradiance at different locations and times have spatial and temporal dependencies.

Also, from the perspective of power demand, loads have dependencies. An example is shown in Fig. 1.1, which was the active loads of five nodes in the French power system in 2012 [4]. The diagonal histograms and scatter plots demonstrate highly non-standard marginal distributions and non-linear correlations of load data. Moreover, load profiles vary with the application of new technologies and policies. With the popularity of EVs (electric vehicles), research has been conducted to orchestrate car charging behaviors [5]; different electricity tariffs are offered to impact customers' electricity usage [6]. These demand side response strategies enhance the flexibility of the power system, but they also lead to changes in the load profile.

1.1.2. RESEARCH FOCUS: ANOMALY DETECTION AND DATA GENERATION

If we look at power system measurements, whether they are time series or snapshots, they generally have dependencies. This dependency means that if you know one or more values, then you can infer information about others. Specifically, time series data represent a series of data points indexed in chronological order, while snapshots indicate data points at particular moments in time. A collection of many such time series and snapshots can be considered a distribution. In an extensive power system, measurements

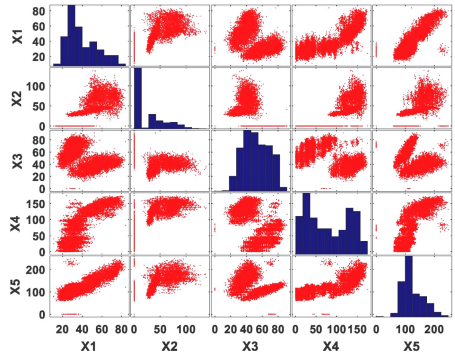

Figure 1.1: Marginals and bivariate distributions of active loads (MW) for five randomly selected buses in France in 2012 [4].

from large amounts of sensors generate multi-dimensional dependent data points, and the collection of those data points further constitutes a multi-dimensional data space with a specific data distribution. Although higher numbers of measurements enable smarter grid operations, high-dimensional stochastic variables with complex univariate and multivariate distributions could also complicate tasks in modeling power system data. For some power system tasks, it is critical to model the distribution of measurements. Two examples are power system anomaly detection and synthetic data generation. These two tasks are interlinked due to the common ground of first capturing the distribution of the historical data set. Concretely, the anomaly detectors can be considered discriminative models, which are vital for power system stability and economic dispatching [7], [8]. Data generation is based on generative models and is essential for system performance assessment [9], [10] when data are insufficient for specific applications. More detailed information on anomaly detection and data generation in the research is elaborated as follows.

PART I: POWER SYSTEM ANOMALY DETECTION

Definition of Anomaly. Among different definitions of the anomaly, a widely accepted one is that "*anomalies are patterns in data that do not conform to a well-defined notion of normal behavior. Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior* [11]." The definitions of anomalies and outliers are sometimes interchangeable. Therefore, another widely accepted definition of anomalies is based on the word "*outlier*." It is defined by Hawkins as follows: "*An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it*

1

was generated by a different mechanism." [12] The '*degree of deviation*' needs to be quantified based on modeling '*other observations,*' but the definition of *other observations* is not explicit, especially for collective outliers [13]. An example is shown in Fig. 1.2. The two-dimensional data set has two normal areas, N_1 and N_2. Intuitively, points that do not lie in those regions, such as a_1 and a_2, are anomalies. However, according to the definition, a_1 and a_2 do not 'deviate significantly from other observations' a_3 - a_6, so they are considered normal. This conclusion may be contrary to intuition. In this research, we prefer to use the first definition that an anomaly is defined as an observation that does not match the patterns inferred from data that are considered normal.

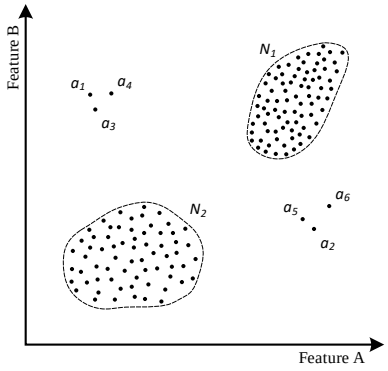

Figure 1.2: A simple example of anomalies in a two-dimensional data set.

Motivating Examples. There have been examples of supervisory control and data acquisition (SCADA) network disruptions caused by cyber attacks. It is known that specific malware, targeting industrial control systems (ICS) and using detailed knowledge of computer systems and programmable logic controllers (PLC), can carry out deliberate attacks to manipulate sensor measurements, application servers (e.g., energy management systems (EMS)) in SCADA systems without being detected (*Stuxnet* [14]). In addition, more recently, the world's first known Hacker-caused blackout happened in Ukraine on December 23rd, 2015. The hackers invaded three energy distribution companies' information systems and shut down 30 substations. This event resulted in more than 200,000 customers living without electricity for 1-6 hours [15]. To learn system knowledge and environment, the attackers performed a long-term reconnaissance of the target network and ultimately conducted multi-site attacks with high synchronization [16]. Those abundant resources, extensive knowledge, elaborate malware design,

and attack experience contribute to the feasibility of conducting well-designed data attacks by corrupting measurements in a coordinated manner as well as keeping stealthy.

Power System Anomalies. Anomaly detection is of considerable significance for secure power system operation and control as well as economic dispatch. This thesis focuses on detecting anomalous measurements, particularly the ones that physically make sense but represent uncommon states. The criteria for "physically make sense" include, but are not limited to 1) Satisfy the design constraints of the power system; 2) Conform to Kirchhoff's law of voltage and current. Such anomalous system states are plausible on technical grounds and do not easily arouse the suspicion of the system operators. Power system anomalies can be broadly classified into two categories: 1) intentional anomalies, such as deliberate attacks, and 2) unintentional anomalies, such as instrument malfunction. Specifically, in the research, we aim to detect intentional anomalies of stealthy power system data integrity attacks [17]. Efforts will also be paid to detect more general unintentional anomalies in renewable energy systems, such as mild reductions in the power output of wind turbines. This may reveal the potential wind turbine component malfunctions. Accurate detection of the above anomalies depends on the precise modeling of historical data sets.

PART II: POWER SYSTEM DATA GENERATION

As mentioned at the beginning of Section 1.1.2, data generation (synthesis or augmentation), another part of this thesis, is also a task that is based on gaining insight into available data. For modern power systems, indeed, more meters and higher measuring frequency result in higher-dimensional data sets with a larger volume. However, sometimes the available data set or the data that researchers are interested in is still too small for desired applications, such as power system risk assessment [18], security analysis [19], and system planning [20]. These applications need abundant but nonrepeating scenarios to comprehensively evaluate system performance, and then gain insight for the next step of power system operation and investment. Moreover, generating data with specific properties is sometimes computationally expensive, especially for properties that are rare, such as in risk assessment applications [21]. In this view, a generative model is needed to efficiently synthesize data with targeted labels. The sparse historical data set is interpolated by a synthetic set to form a new data set with better coverage in multi-dimensional data space. This enriches the data scenarios and could also help to solve data insufficiency challenges of downstream analytical tasks or machine learning-

based applications.

Compared with anomaly detectors that draw boundaries in a data set and predict labels (non-/anomalous) of the data, data-driven data generators try to model how data is placed throughout the space and then generate data on that basis. From the perspective of applications, anomaly detection is used to recognize *existing* abnormal phenomena, while the generative models can be used in synthesizing representative system scenarios and then simulating *potential* power system risks.

In power system data generation research, efforts have been made to generate load profiles for electricity theft detection [22] and electric vehicle charging behavior modeling [23]. This thesis explores the generation of load profiles to further estimate the risk of power systems, which will benefit system operations. This involves the consideration of generator, transmission network capacities, and the negative impact caused by high load demands and limited capacities.

1.2. CHALLENGES AND RESEARCH QUESTIONS

This thesis focuses on anomaly detection and synthetic data generation for power systems. The specific challenges and corresponding research questions are elaborated in this section.

Data is useful and happens to have complex and high-dimensional dependencies. This enables more intelligent grid operations, which, in turn, widen the attack surface and increase the possibility of keeping stealthy [24]. The attackers can define a perturbation that is applied to one or a few measurements of a high-dimensional data vector, actively making the attack be covered up in high-dimensional space and thus resulting in the model misclassifying the perturbed data. The detection of such anomalies requires not only accurately capturing the features of historical data, but also addressing the curse of dimensionality [25]. For the anomaly detection algorithm based on probability and statistics, such as the Gaussian Mixture [26] and Histogram models [27], when the high-dimensional data does not directly meet a specific distribution, the data needs to be divided into different blocks and regarded as a mixture of multiple distributions. However, with the increase of dimension, it becomes more and more difficult to divide the data, and will be easy to overfit, making these algorithms less applicable to high dimensional data sets. Moreover, the widely used distance or density-based detection algorithms, for example, *K-nearest neighbors* (*K-NN*) [28] and *Local Outlier Factor* (*LOF*)

[29], need to calculate the proximity between the given point and other points in the data set. However, in the high-dimensional data set, points become sparse, and manipulated measurements may be masked by the noise of irrelevant dimensions [25]. And for a large volume data set, the proximity calculation between a data point and other points could be another challenge to computational efficiency. In addition, the infrequent occurrence of anomalies in the historical data makes it difficult to explicitly model the pattern of anomalies. Even though the patterns of anomalies are well modeled, the detector only learns to detect known anomalies, which is a significant weakness in novel anomalies. Thus, a feasible detector that is suitable for detecting variable anomalies using an unbalanced data set is needed. Thus, firstly, this thesis focuses on answering the following questions:

Q1 **Power system anomaly detection:** *How to detect anomalies in power systems, especially for anomalies that physically make sense but are uncommon, such as data attacks? How to deal with unbalanced power system data sets due to the rarity of anomaly data? How to tackle the challenges arising from novel anomaly patterns?*

After investigating the research question proposed in **Q1**, the autoencoder (AE) neural network [24] is identified as a promising framework for anomaly detection. With the goal of achieving minimal average reconstruction error, the autoencoder network is trained to replicate inputs at the output side. An anomaly detection boundary is then set according to the distribution of residuals generated from the replication process. However, power system measurements are spatially and temporally dependent. For example, solar or wind power generation within a given region is dependent because of the spatially correlated solar irradiance and wind speed. The dependencies of power system measurements at the input side may also lead to correlated residuals. Nevertheless, when calculating reconstruction errors, traditional Euclidean norm-based calculation schemes [24], [30] don't consider this correlation, which may result in the anomaly detection boundary not fitting the distribution of the correlated residuals. Apart from detecting anomalies, the localization of anomalous data is also of great significance for follow-up anomaly isolation and recovery. In view of this, this thesis also focuses on bridging the gap by answering the following questions:

Q2 **Anomaly detector Enhancement:** *How to deal with the challenge of highly correlated stochastic data when detecting anomalies? Is it possible to enhance the*

1

anomaly detection and localization performance of the AE-based detector proposed in the solution of **Q1** *by taking the correlation of the data into account? What metrics can be used for anomaly localization? Are the anomaly detection and localization improvement strategies the same?*

Anomaly detection is applied to identify abnormal data, while data generation can be used to synthesize data and then analyze potential risks based on the synthesized data. The former can be considered as a discriminative task, i.e., calculating the conditional probability that a given measurement will occur and then determining whether the data is normal or abnormal. The latter is a generative model that considers the joint probability of the measurements. Mathematically, for a given data set X and its corresponding labels Y (non-/anomalous), the discriminative model is trained to achieve a satisfactory conditional probability $P(Y|X)$ (i.e., predict the labels Y with the high true positive and low false positive rate when given X). In contrast, the generative model focuses on generating a new data set that follows the probability of $P(X)$ itself. Both applications essentially require accurate modeling of historical power system data X. For power system risk assessment, it is essential to utilize abundant representative scenarios to assess system performance [19], [20]. Data generation is necessary when data is insufficient for specific applications. However, historical measurements in the data set X are usually high-dimensional. The challenge of generating data based on X is that a generator needs to capture both marginal distributions and multi-variate dependencies of the historical data. The commonly used *Gaussian Copula* model is constructed from a multivariate Gaussian distribution by using the probability integral transform [31]. This model focuses more on fitting the marginal distribution of generative data to historical data. On the contrary, the *Generative adversarial network* (GAN) [32] is powerful in generating data with similar multivariate dependencies to historical data. Moreover, the generated data are new but should be similar to the historical data. The quality of the generations depends on the measuring of this similarity. Visual comparisons can be used for quality validation when data are in the form of figures (e.g., face photos). However, this is not intuitive for power system data in the non-pictorial form. In addition, some power system data have very limited diversity and variability, such as country-level load data, whereas load profiles of individual customers are more stochastic. Generating representative load profiles from a large number of individual users can be more challenging. More importantly, the practicality of generative models should be validated in near-real

case studies. This thesis aims to address the above challenges by answering the following research question:

Q3 Power system data generation: *How to generate power system data by considering both univariate distribution and multivariate correlations? How to evaluate the quality of the generated data? How can the generated data be used practically in risk assessment tasks? Is the proposed generative model capable of synthesizing data with different aggregation levels?*

By answering research questions proposed in **Q3**, a (conditional) variational autoencoder ((C)VAE) is demonstrated to be a data generator that embeds both marginal data distribution and multi-variate data dependencies. Compared with an autoencoder neural network, the inputs of a (C)VAE are compressed to latent space codes with an extra constraint that the codes follow a specific distribution. However, from the perspective of users, not just a data generator is needed, but also a generator that can create data with specific characteristics. Ideally, these characteristics can be inferred from the latent space code. Then, the sampling process will be more targeted, and the data generation process will be more controllable. However, for the (C)VAE model, the location of codes that correspond to specific properties is not known. This would not enable a generation process using user-defined goals. On top of that, the generation of data with specific characteristics requires data with corresponding labels (related to target characteristics) fed into the generative model for training. However, the analysis of some labels is time-consuming. For example, labeling a power system state with a particular adequacy metric, Loss Of Load Expectation (LOLE [h/year]) [33], involves extensive simulations. Thus, only a small proportion of data sets may be labeled. In order to lay the groundwork for training generators using data sets incompletely labeled with hard-to-get information, it may be necessary to validate the feasibility of training generative models in a semi-supervised manner. Therefore, as the last part of this thesis, it focuses on addressing the following challenges:

Q4 Controllable data generator: *Based on the generator proposed for tackling* **Q3**, *how to improve the VAE-based generative model to make the data generation a controllable process that synthetic data with specific user-defined goals? Is it possible to train a generative model with an incompletely labeled data set?*

1

1.3. CONTRIBUTIONS AND THESIS OUTLINE

As the first part of this thesis, Chapter 2 and Chapter 3 develop anomaly detection approaches. Chapter 4 and Chapter 5, the second part of this thesis, focus on providing a generative model to synthesize snapshots of load states and time-series load profiles. The structure of this thesis is depicted in Fig.1.3. The outline of the research work is as follows. Chapter 2 proposes an autoencoder-based anomaly detector to detect false data injection attacks where the labeled anomalous data is insufficient for supervised learning methods. To enhance the detection and localization performance, Chapter 3 moves one step further to revise the model proposed in chapter 2 by using whitening transformation schemes. In Chapter 4, a conditional variational autoencoder (CVAE) is implemented to generate both country-level load states and individual customers' load profiles. It is a variation of the autoencoder neural network with an extra constraint of latent sampling. On the basis of Chapter 4, Chapter 5 proposes an oriented variational autoencoder-based generative model to make the data generation process more controllable. Eventually, Chapter 6 concludes the thesis and proposes future research directions.

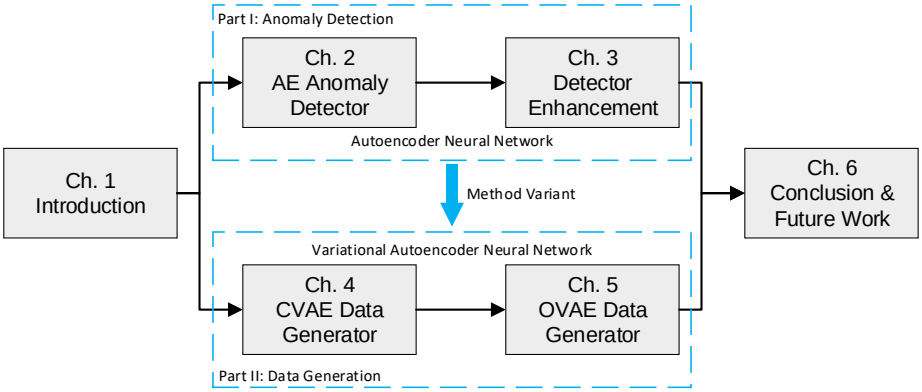

Figure 1.3: Outline of this thesis.

The contributions of this thesis consist of the detection algorithm implementation, novel data processing scheme, detection and data generation quantifying metrics, and new generative model design, which are summarized as follows:

- **Chapter 2** This chapter contributes to addressing **Q1** by detecting power system data attacks in high dimensional space using a highly unbalanced historical data

set. It proposes an autoencoder neural network-based anomaly detector to reduce the dimension of power system measurements as well as learn their patterns. The detector considers anomaly detection as a one-class classification task by acquiring the patterns of enormous normal historical operation states only. Rare anomalies which deviate from patterns learned from normal states are then detected. By only focusing on what is normal, it could be prepared for novel data attacks conducted by resourceful attackers. To guide other researchers to implement an autoencoder-based anomaly detector, this chapter also investigates the influence of the hyperparameter selection on the training and anomaly detection performance.

- **Chapter 3** To answer **Q2**, this chapter extends the work of Chapter 2 to improve the detection performance of the autoencoder-based anomaly detector and start investigating the anomaly localization performance. The detection and localization performance is demonstrated using a case study of renewable energy generation patterns. The anomaly detection model is revised by implementing novel data whitening transformation scheme for input data and/or residuals. Three anomaly localization metrics are proposed to quantify the dependability of the anomaly detector, measuring the degree of standout of anomalous dimensions and their difference from normal dimensions. It is shown that different whitening matrices and whitening timing vary the detection and localization performance. Based on quantitative experimental results, it is found to be most beneficial to use whitening transformations in both input and residual processing stages to enhance anomaly detection performance. For anomaly localization, only whitening data in the latter stage is recommended.

- **Chapter 4** The work of this chapter shows how to implement a conditional variational autoencoder (CVAE) neural network-based generative model to generate multivariate load states and profiles. This CVAE implementation goes beyond common implementation by including stochastic variations of the output samples and co-optimizing the parameters for the output variability. Subsequently, this chapter continues to tackle **Q3** by proposing to use 3 quantifying metrics to statistically evaluate the quality of generated country-level load states. The evaluation results show a satisfactory generation capacity of the CVAE-based model to embody both univariate distribution and multivariate correlations. A simple

multi-area adequacy assessment model is introduced to demonstrate the practicability of the generative model to assess power system risk. After laying this basis, this chapter proposes 3 data processing strategies that are useful when synthesizing load profiles of individual electricity customers.

- **Chapter 5** On the basis of chapter 4, this chapter continues to solve the questions proposed in **Q4**. A novel oriented variational autoencoder (OVAE)-based generative model is proposed to correlate the latent space code and user-analyzed labels. This would result in a more controllable load state sampling process compared to the naive VAE-based generative model. Additionally, the targeted data sampling process is sped up with well-defined importance weights. The effectiveness of the OVAE model is evaluated at both training and generation stages in visual and statistical manners. Moreover, in order to lay the groundwork for dealing with situations where the analysis of certain data labels is time-consuming, e.g., risk labels of load states analyzed by Monte Carlo simulations, a semi-supervised learning scheme is proposed to train the OVAE model with incomplete labels. It is shown that even trained with incomplete data labels, the generative model is able to speed up the targeted data sampling process with well-specified importance weights.

The contributions of this thesis are completely based on papers published and submitted during my Ph.D. research. The papers related to each part of this thesis are listed as follows:

Anomaly Detection

C. Wang, S. Tindemans, K. Pan, and P. Palensky, "Detection of false data injection attacks using the autoencoder approach", in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, IEEE, Liege, Belgium, 2020, pp. 1–6. DOI: 10.1109/PMAPS47429.2020.9183526

C. Wang, K. Pan, S. Tindemans, and P. Palensky, "Training strategies for autoencoder-based detection of false data injection attacks", in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, Den Haag, the Netherlands, 2020, pp. 1–5. DOI: 10.1109/ISGT-Europe47291.2020.9248894

C. Wang, S. Tindemans, and P. Palensky, "Improved Anomaly Detection and Localization Using Whitening-Enhanced Autoencoders", *IEEE Transactions on Indus-*

trial Informatics, Accepted. DOI: 10.1109/TII.2023.3268685

Data Generation

C. Wang, E. Sharifnia, Z. Gao, S. H. Tindemans, and P. Palensky, "Generating multivariate load states using a conditional variational autoencoder", *presented* in *XXII Power Systems Computation Conference (PSCC 2022)*, Porto, Portugal, 2022 and *published* in *Electric Power Systems Research*, vol. 213, p. 108603, 2022.

C. Wang, S. H. Tindemans, and P. Palensky, "Generating contextual load profiles using a conditional variational autoencoder", in *2022 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, Novi Sad, Serbia, 2022, pp. 1–6. DOI: 10.1109/ISGT-Europe54678.2022.9960309

C. Wang, E. Sharifnia, S. Tindemans, and P. Palensky, "Targeted Analysis of High-risk States Using an Oriented Variational Autoencoder", *IEEE Transactions on Power System, Submitted.*

2

Autoencoder-based Anomaly Detection

State estimation is of considerable significance for the power system operation and control. Well-designed false data injection attacks can utilize blind spots in conventional residual-based bad data detection methods to manipulate measurements in a coordinated manner and thus affect the secure operation and economic dispatch of grids. In this chapter, we propose a detection approach based on an autoencoder neural network. By training the network on the dependencies intrinsic in 'normal' operation data, it effectively overcomes the challenge of unbalanced training data that is inherent in power system attack detection. To evaluate the detection performance of the proposed mechanism, a series of experiments on the IEEE 118-bus power system are conducted. Further, the impact of hy-

This chapter is based on the following published work:

[24] C. Wang, S. Tindemans, K. Pan, and P. Palensky, "Detection of false data injection attacks using the autoencoder approach", in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, IEEE, Liege, Belgium, 2020, pp. 1–6. DOI: 10.1109/PMAPS47429.2020.9183526

[30] C. Wang, K. Pan, S. Tindemans, and P. Palensky, "Training strategies for autoencoder-based detection of false data injection attacks", in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, Den Haag, the Netherlands, 2020, pp. 1–5. DOI: 10.1109/ISGT-Europe47291.2020.9248894

The conventional bad data detector in Section 2.4.2 was developed by Kaikai Pan.

perparameters on the detection performance for false data injection attacks that target power flows is investigated. The experiments demonstrate that the proposed autoencoder detector displays robust detection performance under a variety of attack scenarios.

2.1. INTRODUCTION

2.1.1. STATE OF THE ART OF ANOMALY DETECTION

The power system is increasingly equipped with sensors and communication infrastructures. This enables smarter grid operations, but also increases the likelihood of inaccurate measurements. More worryingly, it also enables novel cyber attack scenarios that manipulate power system measurements instead of directly disrupting information and communication technology (ICT) infrastructure or stealing valuable data. Although the typical bad data detection (BDD) within state estimation (SE) can detect erroneous measurements and some basic attacks, well-designed attacks can remain stealthy and bypass the BDD, such as the stealthy false data injection attacks (FDIAs) [34]. These stealthy measurement manipulation attacks severely threaten both the economic dispatching and security control of the power system [7], [8].

There are several common universal anomaly detection methods, which can be divided into a few categories: probabilistic and statistical-based, proximity-based, and dimension reduction and reconstruction-based [13]. They are elaborated as follows.

Probabilistic and statistical-based: The basic idea of the probabilistic and statistical-based anomaly detection methods is assuming the data follows a given parametric distribution, such as the Gaussian distribution, and then calculate the occurrence probability of the measurements. If the probability is lower than the set threshold, the data point is considered an anomaly.

Proximity-based: Anomaly detection methods based on proximity generally belong to 3 categories: distance-based, density-based, and cluster-based. These methods are similar and related to each other with small differences. The definitions of these methods are briefly introduced below.

- **Distance-based:** Distance-based anomaly detection measures the distance between the given data point and its adjacent data points. An example of a distance-based anomaly detection method is the K-nearest neighbors (K-NN) algorithm [28]. Concretely, the distance of the given point to its k^{th} nearest neighbor is regarded as the anomaly score. If the distance (anomaly score) is above a threshold, the data point is considered an anomaly.

- **Density-based:** Density-based anomaly detection defines an anomaly as data that lies in the sparse region. The amount of data points in a specific region is used

to define the density, which can be converted to the anomaly score. If the anomaly score (density) is higher (lower) than a certain threshold, the given data point is an anomaly. The Local Outlier Factor (LOF) [29] is a commonly used density-based anomaly detector.

- **Cluster-based:** For clustering-based anomaly detection, data is first clustered. Then, if a data point does not belong to any clusters or belongs to a small cluster, this data point can be considered an anomaly. The well-known K-means clustering [35] is a representative algorithm that observes data points and clusters similar ones into a predetermined number of K groups. However, it is worth noting that the major difference between density and clustering-based methods is that density-based methods segment the data space, whereas clustering methods segment the data points.

Dimension reduction and reconstruction-based: The dimension reduction and reconstruction-based anomaly detection is a method of extracting features to the latent space with lower dimensions and then reconstructing latent space codes to the original high-dimensional data (the latter can be omitted). There are two fundamental ways to reduce the dimensions, i.e., linear-based and non-linear-based. The corresponding typical algorithms are principal component analysis (PCA) [36] and autoencoder [37], respectively. The difference between the reconstructed and original data is denoted as the reconstruction error, which can then be used to determine the detection threshold. Ultimately, data points with reconstruction errors higher than the threshold are considered anomalies.

Based on different categories of universal anomaly detection principles, several techniques have been proposed to deal with stealthy FDIAs. In [38], the authors have proposed a Kalman filter estimator together with a chi-square detector. Other statistical methods, such as sequential detection using Cumulative Sum (CUSUM)-type algorithms, were designed in [39]. The recent work [40] has proposed a detector utilizing the statistical consistency of measurements, presuming that the system is observable by a minimal set of secure phasor measurement units. These methods, however, can be limited by the prior assumption that measurements fit specific distributions, or by restrictions on the number of manipulated measurements [41].

Moreover, it is increasingly recognized that the distribution of normal power system states is not easily characterized using standard parametric distributions [4]. The need to

operate in a complex stochastic environment has led to the deployment of data-driven methods. Distance-based algorithms like k-NN were used to cluster normal and corrupted measurement states [42]. Nevertheless, the very high dimensionality of measurements (from the physical, cyber, and market domains) results in data sparsity, where manipulated measurements may be masked by the noise of multiple irrelevant dimensions. This can make detection using a high-dimensional distance-based algorithm computationally inefficient or even invalid [25].

Alternative data-driven approaches to FDIA detection have been proposed in the form of support vector machine (SVM)-based classifiers [43] and deep neural network-based classifiers [44]. Both are supervised machine learning algorithms that classify measurements into normal and manipulated data on the basis of labeled training data. However, due to the infrequent occurrence (or more likely: absence) of FDIAs in historical data, the training data set is highly unbalanced, so that it must be augmented by simulated training data. Moreover, in this way, the detector only learns to detect known attacks, which is a significant weakness in a fast-evolving field with resourceful and potentially well-equipped attackers.

2.1.2. CONTRIBUTION AND OUTLINE

This chapter bridges the identified gap by proposing a detection approach based on an autoencoder neural network. The main contributions of this chapter are listed below:

1) We propose an autoencoder-based detection approach for FDIAs. It learns to identify anomalous system states (and therefore possible attacks) using only SCADA-type power flow measurements for a large range of normal operating conditions. Therefore it is well-suited to the inherent data imbalance in FDIA detection applications.

2) We define a case study on the IEEE 118-bus system, including a model to generate 'normal' data. We formulate two FDIA scenarios by considering comprehensive factors of the adversaries' purpose, capacity, and knowledge and utilize indicators to evaluate the FDIA detection performance of our proposed mechanism. The experimental results demonstrate the mechanism has satisfactory detection accuracy.

3) We describe an autoencoder-based detection approach for FDIAs and investigate

the influence of the hyperparameter selection on the training and FDIA detection performance of the proposed mechanism. Experimental results show that the mechanism has the ability to achieve good learning efficiency and detection accuracy.

The outline of this chapter is as follows. Section 2.2 briefly reviews the state estimation and the bad data detection technique, followed by the formulation of the false data injection problem. Section 2.3 proposes an FDIA detection mechanism based on the autoencoder neural network. The attack detection theory, training, and detection processes of the detector are elaborated. Section 2.4 utilizes the IEEE 118-bus system to assess the detection performance of the proposed autoencoder-based attack detector in load-targeted attack scenarios for economic profit. Section 2.5 investigates the influence of hyperparameters on the training and detection performance of the detector.

2.2. ANOMALOUS DATA ATTACK SCENARIOS

A well-designed false data injection attack can bypass the traditional bad data detection scheme and mislead the state estimate process. To formulate the false data injection problem, the background information of the power system state estimation process and bad data detection techniques are introduced.

2.2.1. POWER SYSTEM STATE ESTIMATION

The power system we consider has n_b buses and n_t transmission lines. The vector $\theta = [\theta_1, \theta_2, \ldots, \theta_{n_b}]^T$ represents n_b phase angles, excluding the angle of the reference bus. In this section, a DC power flow model is assumed, in which the reactive power is neglected and bus voltages are assumed to be 1 (p.u.). The vector $P^I \in \mathbb{R}^{n_b}$ of active power injections is related to the angle vector θ,

$$P^I = AP^F = AR^{-1}A^T\theta, \tag{2.1}$$

where $P^F \in \mathbb{R}^{n_t}$ is the branch active power flow vector, $R \in \mathbb{R}^{n_t \times n_t}$ is a diagonal matrix of transmission line reactance and $A \in \mathbb{R}^{n_b \times n_t}$ is the branch-to-node incidence matrix [45]. In the following, we use the power injection vector P^I as the system state $x \in \mathbb{R}^{n_b}$. It is functionally equivalent to the more commonly used phase angle vector θ, but it allows for the more elegant generation and detection of FDIAs.

We consider a system where the active power injections and line flows are measured with some error. Thus the system model $H \in \mathbb{R}^{(n_b+n_t)\times n_b}$ for measurement and state can be written by

$$z = \begin{bmatrix} I \\ H^F \end{bmatrix} x + e = Hx + e, \tag{2.2}$$

where the measurement noise vector $e \sim \mathcal{N}(0, D)$ denotes m=n_b+n_t independent zero-mean Gaussian variables with the covariance matrix $D = \text{diag}(\delta_1^2, \ldots, \delta_m^2)$ (measurement noise are assumed to be independent). The measurement vector $z \in \mathbb{R}^m$ indicates measured power injection and line power flow with noise. Identity matrix $I \in \mathbb{R}^{n_b\times n_b}$ and distribution factor matrix $H^F \in \mathbb{R}^{n_t\times n_b}$ are parts in H corresponding to the power injection and line power flow, respectively. According to (2.1), the distribution factor matrix can be described as $H^F = R^{-1}A^T(AR^{-1}A^T)^{-1}$. Given the observation of the measurements z, the state estimate $\hat{x}$ is solved by the weighted least squares (WLS) approach [46] as

$$\hat{x} = (H^T D^{-1} H)^{-1} H^T D^{-1} z := Kz. \tag{2.3}$$

2.2.2. STEALTH FALSE DATA INJECTION ATTACKS

The vector $\hat{x}$ is then utilized to estimate the power injection and line power flow measurements by $\hat{z} = H\hat{x}$. In bad data detection, a residual is defined to describe the difference between the actual and the estimated measurements, namely $r_o = z - \hat{z}$. This naturally gives rise to a BDD scheme that identifies bad data by comparing the 2-norm of r_o with a certain threshold τ, i.e. an alarm is triggered if $\|r_o\|_2 > \tau$.

We denote $a \in \mathbb{R}^m$ as the non-zero false data vector injected into measurement vector z. The manipulated measurement vector can be described as $z_a = z + a$. Here the vector c is defined as the deviation of the estimated state before and after the attack. The corrupted system state can be denoted as $\hat{x}_a = \hat{x} + c$. According to (2.3), the falsified state estimate $\hat{x}_a$ can be written by

$$\begin{aligned} \hat{x}_a &= (H^T D^{-1} H)^{-1} H^T D^{-1} z_a \\ &= (H^T D^{-1} H)^{-1} H^T D^{-1} (z + a) \\ &= \hat{x} + c, \end{aligned} \tag{2.4}$$

and the corresponding r_a after the attack can be expressed as

$$\begin{aligned} r_a &= z_a - H\hat{x}_a = z + a - H(\hat{x} + c) \\ &= r_o + (a - Hc). \end{aligned} \tag{2.5}$$

If $a = Hc$, then the manipulated residual r_a equals the original residual r_o. Thus the attacker manipulates the measurements with the residual unchanged and keeps stealthy with respect to this physics-based BDD scheme. This remains true if $a \neq Hc$, as long as $\|r_a\|_2 \leq \tau$ is still satisfied.

For our FDIA detection study, we consider attack scenarios from the perspective of an adversary that manipulates load patterns [8], for example in order to hide excessive power consumption or to reduce apparent power consumption for economic motives. The adversary needs to corrupt the power generation and power flow accordingly to avoid detection by BDD. The real state (and therefore ideal estimation) of the power system is $\hat{x}$. After injecting the attack vector a, the measurements and state estimation change to $z+a$ and $\hat{x}+c$, respectively. The attack scenario will be detailed in Section 2.4.

2.3. AUTOENCODER-BASED ANOMALY DETECTORS

To address the challenge of the detection of false data injection attacks with a highly unbalanced data set and the possible variation of novel data attacks, an autoencoder neural network-based detector is proposed. The specific characteristics and advantages of the method for identifying FDIAs in the context of the power system are first analyzed. Then, the attack detection principle of the autoencoder-based mechanism is explained in detail. Finally, the complete training and detection processes of the proposed mechanism are described.

2.3.1. DETECTOR SCHEMATIC

FDIA detection is essentially a classification problem with the objective of distinguishing false data from data that is considered 'normal'. What the SVM-based [43] and deep neural network-based classifiers [44] have in common is to treat FDIA detection as a supervised learning task. However, supervised learning requires a training data set with representative examples of normal system operations and attacks. Such data sets are in short supply, because of the rarity of attacks, unwillingness to share data, and evolving attacks. As a result, it is difficult to learn a satisfactory discriminator of 'normal' and 'attack' scenarios on this basis [47].

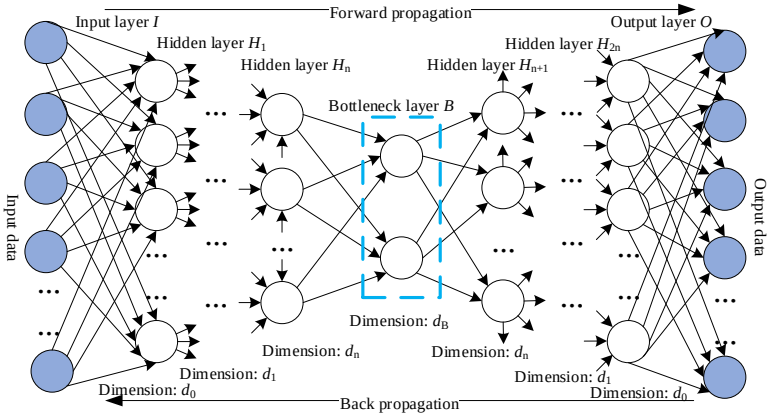

Figure 2.1: The schematic of the Autoencoder.

Instead, we propose to approach FDIA detection as a one-class classification problem, where the detector is trained on examples of only 'normal' operation data. Observations with features that deviate substantially from those in the training data will be considered anomalies, in this case as 'potential attacks'. There are two main advantages to this approach. First, the autoencoder-based mechanism avoids the need to gather or generate attack data to create balanced data sets for training the classifiers. Second, by focusing on what is normal only, the proposed mechanism is naturally prepared for unknown attack patterns.

Autoencoders learn the most important features of the training data (i.e. normal power system measurements) by sending the measurements through an information bottleneck while attempting to reconstruct the training data with minimal error [37]. The structure of the autoencoder algorithm is depicted in Fig. 2.1. The dimension reduction process of mapping the d_0-dimensional input data to the code in the bottleneck layer B through hidden layers H_1 to H_n is named the *encoder*. Afterwards, the *decoder* decompresses the code to d_0-dimensional output data. Weight matrices W and bias vectors b are utilized in the encoding and decoding process as

$$Y = \sigma(W_n^e(\ldots\sigma(W_0^e Z + b_0^e)\ldots) + b_n^e), \tag{2.6a}$$

$$\hat{Z} = \sigma(W_n^d(\ldots\sigma(W_0^d Y + b_0^d)\ldots) + b_n^d), \tag{2.6b}$$

where W_n^e and W_n^d denote weight matrices for encoding and decoding process respectively, b_n^e and b_n^d are bias vectors, and σ represents a nonlinear element-wise activation

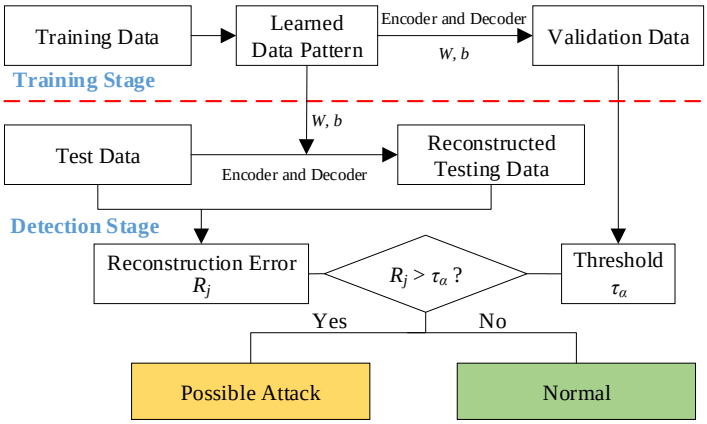

Figure 2.2: The proposed training and FDIA detection mechanism.

function. Z refers to the input data vector, Y is the data in the bottleneck layer and vector $\hat{Z}$ stands for the output data.

2.3.2. ANOMALY DETECTION MECHANISM

The residual associated with a training observation Z_j is given by $r_j = Z_j - \hat{Z}_j$. The reconstruction error R_j is expressed as $\|r_j\|^2/d_0$. The objective of the training process is to minimize the sum of all reconstruction errors R_j as

$$\min_{W,b} \quad \left\{ J := \frac{1}{S} \sum_{j=1}^{S} (\|r_j\|^2/d_0) \right\}, \tag{2.7}$$

where S denotes the total number of observations used for training. By training the autoencoder on training data that is considered normal, it learns to efficiently encode the features of this data in the bottleneck layer B. Data that deviates from the training data in a structural way is therefore highly likely to have a larger reconstruction error.

The training and FDIA detection process of the proposed mechanism is depicted in Fig. 2.2. In the training stage, the algorithm iteratively updates the value of weight matrices W and bias vectors b until the function J converges. At the end of the training process, the reconstruction errors R_j for the *validation* set are sorted in ascending order. A threshold τ_α equals to the α^{th} percentile is then chosen, for example at the value where an 'inflection point' occurs in the error distribution. A possible FDIA is detected when, for a measurement Z_j in the *test* set, the reconstruction error R_j exceeds the threshold τ_α.

2.4. DATA ATTACK DETECTION USING AUTOENCODER

This section focuses on evaluating the detection performance of the proposed mechanism using a case study on the IEEE 118-bus system. First, we describe the process of modeling normal operating conditions and explain how to create anomalous attack scenarios. Then, we describe and analyze the load-targeted attack scenario. For this scenario, we will first quantify the detection performance of our proposed detection mechanism. Specifically, the detection probability, false positive rate, and false negative rate are tested. Next, the detection performance of our detector will be compared with a conventional BDD detector. To do so, we introduce "knowledge limited" attacks that both detectors can potentially detect. Notably, the "knowledge-limited" attacks are more of interest in reality as the attacker may have an inaccurate (e.g. outdated or estimated) system model.

2.4.1. TEST SYSTEM MODELING

MODELING NORMAL OPERATING CONDITIONS

With the long-term secure and stable operation, the power system has a large number of normal operating conditions which involve a significant volume of loads, power generations, and power flows data set. Trained by these data, the proposed mechanism will acquire the data pattern which represents the model of normal system operating conditions.

In the IEEE 118-bus system, electricity is supplied by $M = 54$ generators, transmitted via $Q = 186$ branches and ultimately consumed by $N = 99$ loads. We generated 'normal' (i.e. physically feasible and economically reasonable) power system states and corresponding measurements by using optimal power flow solutions.

Second-order polynomial cost functions were assumed for generators, i.e., $f(P_g^G) = C_{g,2}(P_g^G)^2 + C_{g,1}P_g^G$. Hence the economic dispatch P^{G^*} was solved with the objective to minimize the total generation cost. The solutions were implicitly parameterized by the nodal load P_l^L and generation cost parameter as

$$P^{G^*} = \arg\min_{P^G} \sum_{g=1}^{M} C_{g,2}(P_g^G)^2 + C_{g,1}P_g^G \tag{2.8}$$
$$\text{s.t.} \quad \sum_{g=1}^{M} P_g^G - \sum_{l=1}^{N} P_l^L = 0,$$

where the injection $P^I = P^I(P^G, P^L)$ was determined by the mapping of load P^L and

generation P^G onto the nodes.

Normal operating conditions were generated using a data set that contains a total of 43,717 historical hourly loads from 32 European countries between 2013 and 2017 [48]. These time series were used to generate a 99 load point time series as follows. The national load time series were first divided by 1000, to obtain reasonable magnitudes for individual buses. Then each load point was assigned a random linear combination of the 32 sources by sampling from the Dirichlet distribution with vector valued parameter $(1,\ldots,1)^T$, which generated a uniform distribution on the 31-simplex. Additionally, a normally distributed variation with a standard deviation of ±5% of the measured value was added to each measurement.

An additional source of randomness was created by randomly sampling the generating cost coefficients of the 54 generators as follows. Coefficients $C_{g,2}$ were sampled uniformly in the range [0.085, 0.1225] \$/MW^{2}h and $C_{g,1}$ uniformly in the range [1,5] \$/MWh. These values span the range of generators included in the IEEE 9-bus system supplied with Matpower [49].

The procedure above was used to generate snapshot injections $P^I = P^I(P^{G^*}, P^L)$, which were converted into line flow measurements using $P^F = H^F P^I$. In this investigation, line transmission limits and generator capacities were not enforced, as the focus of this work was on the recognition of load, generation and power flow patterns. This resulted in a 339-dimensional measurement vector for training, containing 99, 54, and 186-dimensional data of loads, power generations, and line power flows, respectively. Independent measurement noise e was added using a truncated Gaussian distribution with zero mean, a standard deviation of 0.33%, and an absolute value less than 1% of the original value [50]. The generated data set $T \in \mathbb{R}^{43717\times 339}$ was divided into a training set $T_r \in \mathbb{R}^{26197\times 339}$, a validation set $T_v \in \mathbb{R}^{8760\times 339}$ and test set $T_e \in \mathbb{R}^{8760\times 339}$ with the ratio 3:1:1.

In this section, the autoencoder network contains 4 hidden layers in the encoder with dimensions of 339, 256, 128, and 64, respectively. The bottleneck layer has 32 nodes, and the decoder maps the 32-dimensional data to a 339-dimensional output through 3 hidden layers with the same dimensions as the encoder. We used the sigmoid activation function between the second and penultimate hidden layer, and the Adam Optimizer [51] was utilized with default settings to iteratively optimize the value of weight matrices W and bias vectors b. The batch size and learning rate for training were 256 and 10^{-5} respectively, and 2,000 training epochs were used. Training and testing of the autoencoder

were conducted using `tensorflow` on the Google Colab environment using the GPU option. Initial performance analysis of hyperparameter settings for the autoencoder-based FDIA detector is available in [52].

CREATING ATTACK SCENARIOS

We developed feasible FDIAs from the perspective of the adversaries by adding an offset to the normal operating conditions created in the previous section. To gain economic profit, attackers inject false data into the grid by using the acquired knowledge of the targeted power system. In the context of this section, this knowledge is represented by the incidence matrix A (topology) and the reactance matrix R of the transmission lines. Moreover, we assumed that the capacity of an attacker was limited by the attackable measurement set [34] and the maximum number of the measurements that the attacker can corrupt simultaneously.

In the following, we quantified the factors described above. According to the attack capacity, the adversary selects a set of attacked loads $\mathscr{L}^A \subseteq \mathscr{L}$. The attacker then determines the modification factor β_l of each selected load and calculates the total load change $\sum_{l\in\mathscr{L}^A} \beta_l P_l^L$, in which $\beta_l P_l^L$ equals the change ΔP_l^L of each load. Similarly, the attack selects a set of attacked generators $\mathscr{G}^A \subseteq \mathscr{G}$. Next, the attacker determines ratios of the power generators' change amount $\lambda_1 : \lambda_2 : \ldots : \lambda_{|\mathscr{G}^A|}$ and normalizes the ratios to get each generator's change ΔP_g^G. Here $|\mathscr{G}^A|$ represents the cardinality of $\mathscr{G}^A$.

$$\Delta P_g^G = \left[\sum_{l\in\mathscr{L}^A} \beta_l P_l^L\right] \times \frac{\lambda_g}{\sum_{g'\in\mathscr{G}^A} \lambda_{g'}} \tag{2.9a}$$

All load changes ΔP_l^L and generation changes ΔP_g^G, together with zeros that denote buses with unchanged injection make up the power injection change vector $\Delta P_A^I \in \mathbb{R}^{118}$. Besides, similar to (2.2), the attacker then utilizes the knowledge of the topology and grid parameters to coordinately calculate power flows change vector $\Delta P_A^F \in \mathbb{R}^{186}$.

$$\Delta P_A^F = H^F \cdot \Delta P_A^I, \tag{2.9b}$$

Afterwards, the attack vector a consists of the change vector of loads, power generations and line power flows.

The FDIA manipulates the original data of loads, power generations, and line power flows. The pattern of the corrupted data may deviate from that of normal operating conditions, which enables it to be detected by the autoencoder if the reconstruction error R_j exceeds τ_α.

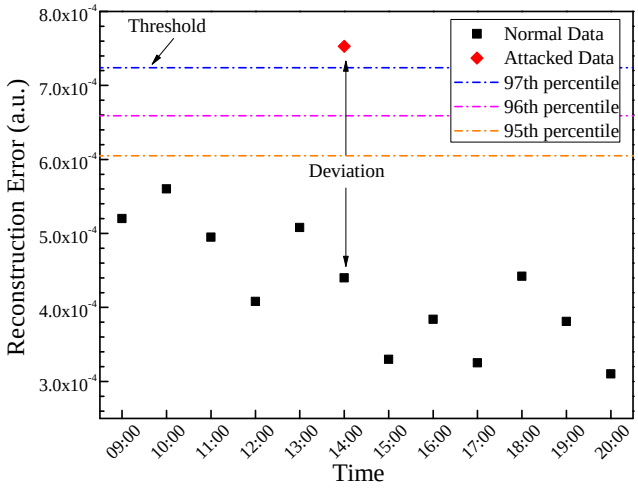

Figure 2.3: Detection effectiveness validation by launching an FDIA.

2.4.2. DETECTION PERFORMANCE ANALYSIS

DETECTION EFFECTIVENESS VALIDATION

We first validated the effectiveness of the trained detector. In this experiment, we observed the change of the reconstruction error R_j before and after a false data injection attack and compare it with the threshold τ_α. A common scenario for an attack happens when the adversary gets the data of a local area and utilizes it to manipulate the neighboring measurements. Here, we selected 12 hours' operating data from 9:00 to 20:00 on December 31st, 2017 as an example. At 14:00, to gain economic profit, an attacker modified three loads profiles of bus 108, 109, and 110, by injecting false data to decrease the power demand of the loads by 10% as -7.48 MW, -5.69 MW and -6.28 MW respectively. Accordingly, to balance the power of loads and generations, the attacker decreased the nearby power injection of two generators connected to bus number 110 and 111 with the ratio $\lambda_1 : \lambda_2 = 1$. Based on (2.9b), the corresponding transmission line power flows were obtained. The experiment result is depicted in Fig. 2.3. From the result, we can observe that before the attack, the reconstruction error R_j of normal operating data is in the range of 3.10×10^{-4} and 5.60×10^{-4}, and they are lower than the threshold $\tau_{97} = 7.25 \times 10^{-4}$ learned in the training process shown in the Section 2.3.2. To be specific, after observing the reconstruction error distribution of the validation data, the threshold is set as 97^{th} percentile due to the occurrence of the 'inflection point' where

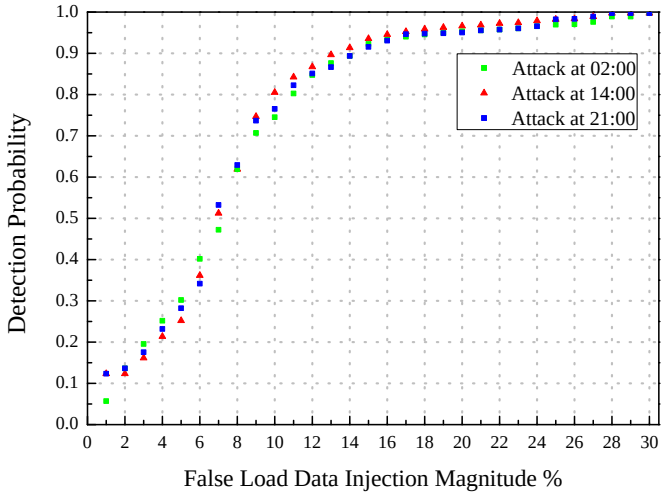

Figure 2.4: Detection probability of attacks at different times with different false load data injection magnitude.

the cumulative distribution curve of the reconstruction error flattens out from the steep rise. After manipulation by the false data injection, the reconstruction error R_j at 14:00 increases from 4.40×10^{-4} to 7.53×10^{-4}, which exceeds the threshold τ_{97} and triggers an alarm. The detector thus recognizes an anomaly in the corrupted measurements, which deviate from measurements taken in normal operating conditions. This result demonstrates that the autoencoder is capable of FDIA detection in at least some scenarios.

GENERAL DETECTION PERFORMANCE

In addition to the one-off effectiveness demonstrated above, we were also interested in its statistical detection performance. This was tested by launching a larger number of FDIAs at various times and with various false load data injection magnitudes. Here the magnitude was defined as the percentage of load reduction in targeted nodes. For the sake of comparison, the attack targets remained the same as those utilized in the last experiment. In this experiment, we launched an attack at 2:00, 14:00, and 21:00 in each day of 2017 by reducing reported loads between 1% to 30% and observing the detection performance. The detection probability is the ratio of detected attacks to all the launched attacks, namely the true positive rate. The results are shown in Fig. 2.4.

Because the load demands at 2:00, 14:00, and 21:00 differ significantly, the resulting power system states (including flows) are also substantially different. However, the result shows, under the same false load injection magnitude, the detection probabilities differ

Table 2.1: Detection performance evaluation.

	Normal Data		Attack Data
True Negative	96.5% (8453)	True Positive	93.6% (8199)
False Positive	3.5% (307)	False Negetive	6.4% (561)

only slightly. This demonstrates that the autoencoder learns the intrinsic relationship of the loads, power generations and power flows from different operating conditions, leading to robust detection results.

In addition, we launched 8760 attacks, one for each hour of 2017, by decreasing the power demand of the same buses by 15%. Besides, we used the hourly normal operating data in 2017 as a control group. The result is shown in Table 2.1.

From the experiment result, we can find that the detection probability (true positive rate) is 93.6%, which denotes a satisfactory detection performance. As mentioned in the first experiment, the threshold τ_{97} was used, corresponding to a 3% misclassification rate in the validation set. It is worth noting that the false positive rate is comparable to the 3.5% observed in Table 2.1. This result suggests that the autoencoder has a good generalization capability and does not overfit.

DETECTION PERFORMANCE COMPARISON

In the above experiments, our proposed autoencoder-based detector has succeeded in generating a diagnosis signal in the presence of FDIAs which can keep stealthy from the viewpoint of BDD. In the second experiment, we compared our detector with BDD in the detection of 'unstealthy' FDIAs. Such attacks have the possibility to be detected by the BDD while the detection ability is intimately related to the topology or parameter errors in the construction of FDIAs by the attacker. Thus in what follows there exist knowledge deviations in the system model acquired by the attacker in computing the attack vector of (2.9). In particular, we explored the case that the attacker knows the exact topology of the network but inaccurate line reactance R in (2.1). This can be described by

$$\hat{R} = R \cdot (I^R + \gamma), \tag{2.10}$$

where $I^R \in \mathbb{R}^{n_t \times n_t}$ is the identity matrix and $\gamma \in \mathbb{R}^{n_t \times n_t}$ is a diagonal matrix whose elements denote the reactance deviation ratio - which we will refer to as the *knowledge deviation ratio*. In this experiment, we ranged the magnitude of the deviations from 0.01

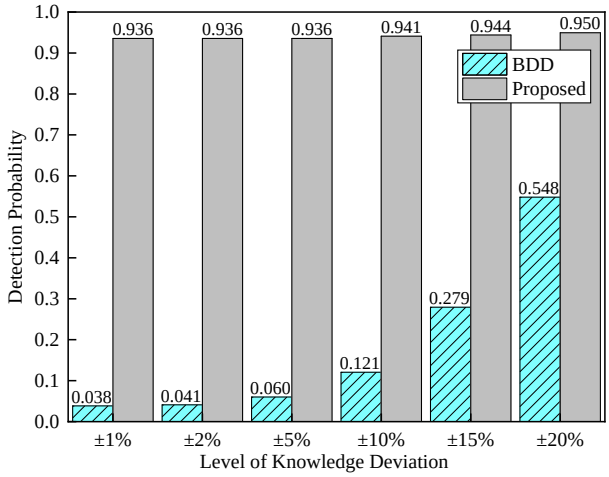

Figure 2.5: Detection performance comparison between the proposed mechanism and the BDD scheme in a load-targeted attack scenario.

to 0.20, with randomly sampled signs for each element. According to the explanation of (2.2), this will lead to an erroneous distribution factor matrix H^F and thus obtain inaccurate power flow values. The BDD scheme usually checks if the weighted p-norm of the measurement residual (called the cost function) is below some threshold τ, which is selected based on the statistical properties of the cost function and a permissible false-alarm rate. In this experiment, 2-norm was used, and the false-alarm rate was set as 5%. Moreover, we kept the attack target unchanged from the previous experiments and set the false load data injection magnitude on the selected three loads by decreasing them by 15%. The results are shown in Fig. 2.5. As the level of knowledge deviation increases from ±1% to ±20%, the detection probability of BDD rises from 0.038 to 0.548, but it remains lower than the detection performance of the autoencoder.

2.5. DETECTOR TRAINING STRATEGY

The performance of a Neural network-based algorithm highly relies on the selection of hyperparameters. In this view, further experiments were conducted on the IEEE 118-bus system to evaluate the influence of hyperparameter selection on the training process and anomaly detection performance.

2.5.1. HYPERPARAMETER TUNING

We study an attack scenario from the perspective of an adversary that aims to interfere with the secure operation of the physical grid by manipulating the power flow measurements. By changing the apparent system state, the attacker can mislead the operator into taking costly or disruptive decisions. The attacker, in general, has limited resources while aiming to stay stealthy from the BDD. In light of this, we consider how many other measurements need to be attacked in coordination with the targeted power flow to avoid triggering alarms. This leads to a constrained optimization problem [53], and the computed optimal value illustrates the minimum number of corrupted measurements in a stealthy attack against the measurement i. This can be written as:

$$\begin{aligned} \min_{a,c} \quad & \|a\|_0 \\ \text{s.t.} \quad & a = Hc,\ a_i = \mu, \\ & a_p = 0,\ \forall p \in \mathscr{P}, \end{aligned} \tag{2.11}$$

where $\|a\|_0$ denotes the number of non-zero elements in attack vector a. Here μ represents the value of injected false data on measurement i. We added the constraint that the measurements in the protected set $\mathscr{P}$ cannot be attacked. It is known that the above optimization program (2.11) is non-convex and may be hard to solve in large problems. However, it can be expressed into a mixed integer linear program (MILP) which can be solved in an appropriate solver with acceptable computation time in an off-line manner.

We conducted experiments on the IEEE-118 bus system. The data processing methods were the same as described in Section 2.4.1. We tuned hyperparameters for the training process by using a grid search over learning rate (10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}) and batch size (64, 128, 256). Other settings of the neural network were the same as in the previous section. The training performance under different parameters combination is shown in Fig. 2.6.

We took a batch size of 256 as an example to illustrate the trend of average reconstruction error $\tilde{r}(z)$ with the increase of training epoch. When the learning rate is set to be 10^{-2}, 10^{-3} and 10^{-4}, the mean value of reconstruction error converges to a high value or exhibits a fluctuation, which indicates high learning rates. However, when the learning rate is 10^{-6}, it makes the convergence error of reconstruction error too slow. Therefore, 10^{-5} is selected as the appropriate value. Near 10^{-5}, we looked for the appropriate learning rate at a higher resolution, and eventually, 10^{-5} and 3×10^{-5}, were set as candidates.

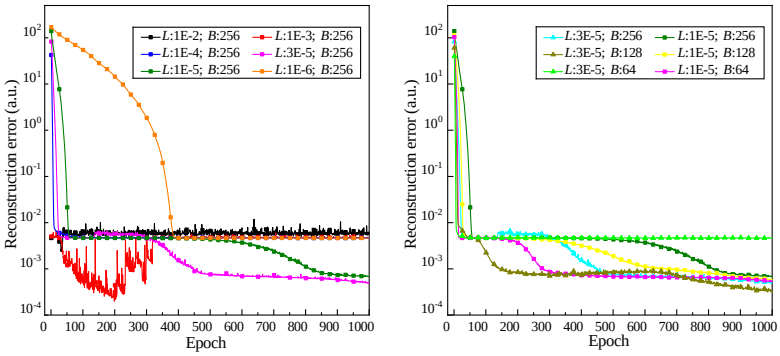

Figure 2.6: The relationship between the training epoch and the reconstruction error. L stands for the learning rate and B represents batch size.

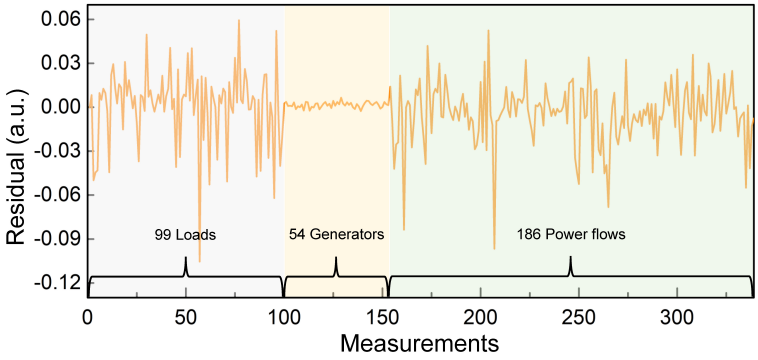

Figure 2.7: Residual of one observation.

Then, we assigned three different batch sizes (64, 128, 256) to the above alternative learning rates and compare the convergence performance. According to the results depicted in Fig. 2.6, in general, a high learning rate and small batch size result in a steeper reconstruction error convergence. In addition, a too-small batch size will increase iterations as well as the training time for running the same training epoch. To make the convergence fast and stable, we selected a learning rate of 3×10^{-5} and a batch size of 256 to train our proposed detector.

Owing to the information loss that happens during encoding and decoding, there exists a residual between the measured data and its reconstruction value. The residuals of one normal observation, which contains 339 measurements, are depicted in Fig. 2.7. Statistical reconstruction errors will be investigated in the following subsections.

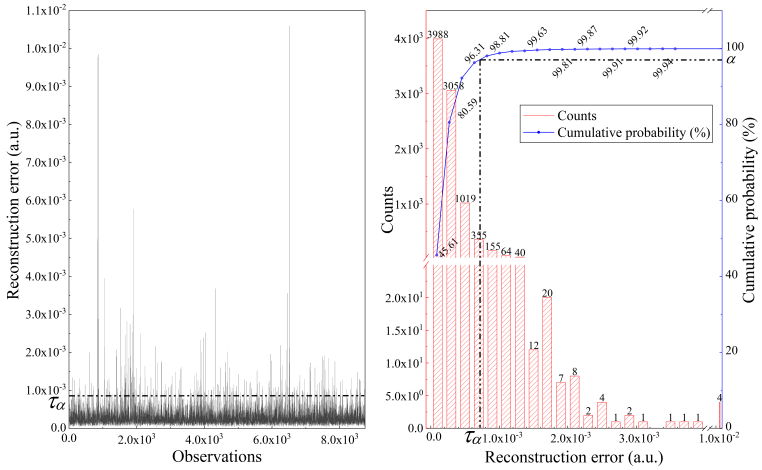

Figure 2.8: Reconstruction errors of the validation data set and the corresponding distribution.

2.5.2. THRESHOLD SELECTION STRATEGY INVESTIGATION

The autoencoder network was trained for 3,000 epochs and the validation set confirmed an absence of overfitting. The reconstruction errors of 8,743 observations were calculated from their residuals by (2.7) and depicted in Fig. 2.8. After sorting R_e in ascending order and observing the their distribution, a threshold τ_α equal to the α^{th} percentile was chosen.

In this study, we selected the branch between bus 109 and 110 to launch power flow-targeted attacks. After solving the MILP of (2.11), the result shows the attacker needs to coordinately manipulate the measured power injection of bus 103, 109, and 110 and the transmission line power flow from bus 103 to 110 at least. We launched 8760 attacks to manipulate hourly observations in the test data set $\mathscr{T}_t \in \mathbb{R}^{8760\times 339}$ by decreasing the power flow from bus 109 to 110 by 10%. Besides, we used the hourly uncorrupted normal operating data in $\mathscr{T}_t$ as a control group. Under power flow-targeted FDIAs, the influence of threshold selection on detection performance is shown in Table 2.2. TP, FN, TN, and FP denote true positive, false negative, true negative, and false positive rate, respectively.

It can be observed that when α is increased from 96 to 100, the false positive rate and true positive rate both decrease. In view of this, α should be set to a sufficiently high value to decrease the false positive rate, but not so high that it comes at the cost of an excessive decrease in the true positive rate. From our experiment, it might be proper to choose an α near the inflection point where the cumulative distribution curve of re-

Table 2.2: The influence of threshold selection on power flow-targeted FDIA detection performance.

	α	τ_α	TP	FN	TN	FP
1	96	7.67×10^{-4}	96.16%	3.84%	92.05%	7.95%
2	97	8.53×10^{-4}	95.62%	4.38%	92.88%	7.12%
3	98	9.89×10^{-4}	92.88%	7.12%	94.25%	5.75%
4	99	1.26×10^{-3}	91.78%	8.22%	96.99%	3.01%
5	99.5	1.67×10^{-3}	89.59%	10.41%	98.08%	1.92%
6	100	1.06×10^{-2}	67.67%	32.33%	100.0%	0.00%

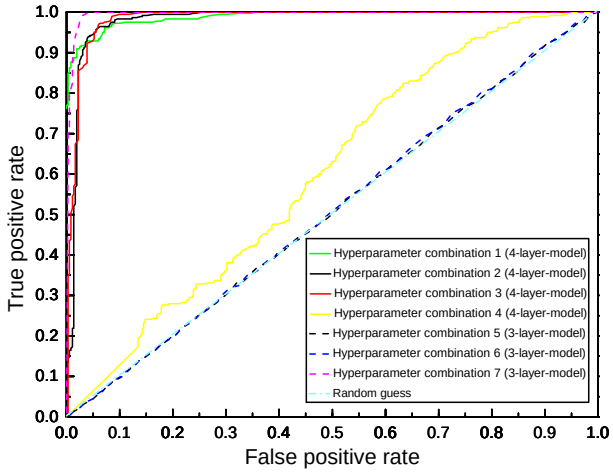

Figure 2.9: Receiver operating characteristic curves

construction errors flattens out from the steep rise. This is consistent with the general practice in anomaly detection. In this case, the value of α is chosen as 99 to give consideration to both more hits (higher true positive rate) and fewer false alarms (lower false positive rate) as 91.78% and 3.01%, respectively.

2.5.3. THRESHOLD SELECTION STRATEGY

In this experiment, we investigated the influence of hyperparameter selection, especially the depth and layer dimension of the proposed model on the FDIA detection performance. We considered 3 and 4-layer models with 7 different dimension configuration combinations as shown in Table 2.3. In 4-layer models, we only changed the dimension of the bottleneck layer. For the 3-layer models, the difference existed in the dimension

Table 2.3: Hyperparameter combination and its FDIAs detection performance. *H*: Hidden layer, *B*: Bottleneck layer

4-hidden-layer models						
	H_1	H_2	H_3	H_4	B	Avg. $\mathscr{R}_t$
1	339	256	128	64	32	2.06×10^{-4}
2	339	256	128	64	24	2.30×10^{-4}
3	339	256	128	64	16	1.99×10^{-4}
4	339	256	128	64	8	4.31×10^{-3}
3-hidden-layer models						
	H_1	H_2	H_3		B	Avg. $\mathscr{R}_t$
5	339	128	64		32	4.68×10^{-3}
6	339	256	64		32	4.67×10^{-3}
7	339	256	128		32	1.90×10^{-4}

combination of the second and third hidden layers. Other training hyperparameters remained the same as in Section 2.5.1 and the attack target remained unchanged from the previous experiments in Section 2.5.2. The result is shown in Fig. 2.9 as receiver operating characteristic curves (ROC) to compare the detection sensitivity (true positive rate) and specificity (false positive rate) under different model configurations.

In 4-layer models, as the dimension of the bottleneck layer decreases from 32 to 16, the models still demonstrate satisfactory detection performance overall. However, if the model is over compressed into the latent space as an 8-dimensional bottleneck layer, it will result in the excessive loss of information during the encoding/decoding process and thus interfere with the detection accuracy. As for the 3-layer-models, the reduction of layers (hyperparameter combinations 5 and 6) may lead to the increase of the reconstruction error which is shown in the last column of Table 2.3 and gives rise to the decline of detection sensitivity and specificity. However, the model with hyperparameter combination 7 still denotes comparable detection probability and false-alarm probability as the 4-layer model. This indicates that setting the dimensions of each layer properly, in particular, selecting wide layers helps to enhance the model's reconstruction and detection capabilities.

3

Anomaly Detector Performance Improvement

Anomaly detection is of considerable significance in engineering applications, such as the monitoring and control of large-scale energy systems. In the chapter, the autoencoder neural network-based anomaly detector developed in chapter 2 is improved to detect more general anomalies in renewable energy scenarios. Specifically, correlations between residuals are identified as a source of misclassifications. In addition, to accurately detect and localize the source of anomalies, whitening transformations that decorrelate input features and/or residuals are analyzed as a potential solution. In this chapter, for a use case of distributed wind power generation, the performance of various data processing combinations is quantified. Whitening of the input data is found to be most beneficial for accurate detection, with a slight benefit for the combined whitening of inputs and residuals. For localization of anomalies, whitening of residuals is preferred, and the best performance is obtained using standardization of the input data and whitening of the residuals using the ZCA or ZCA-cor whitening matrix with a small additional offset.

This chapter is based on the following work:

C. Wang, S. Tindemans, and P. Palensky, "Improved Anomaly Detection and Localization Using Whitening-Enhanced Autoencoders", *IEEE Transactions on Industrial Informatics, Accepted.* DOI: 10.1109/TII.2023.3268685

3.1. INTRODUCTION

3.1.1. RELATED WORK AND MOTIVATION

Monitoring and control of large-scale engineering systems require accurate measurements and dependable communication infrastructure – and methods to process that data for operational awareness. An important example is the case of electrical power systems that are increasingly reliant on variable renewable generation [54]. In this context, it is important to detect anomalies in high-dimensional, highly variable observations from a multitude of sensors. For example, mild reductions in power generation caused by a wind turbine component malfunction or physical disturbance. Insufficient performance of anomaly detectors may threaten both the economic dispatch and secure control of power systems [7].

In recent years, with the development of deep neural network-related technologies, an unsupervised data-driven approach has been proposed in the form of an autoencoder-based classifier [37]. It considers anomaly detection as a one-class classification task by learning patterns of normal operating states. This is well-suited to the inherent data imbalance in anomaly detection applications and the fast-evolving power grid [24]. On this basis, techniques have been designed to detect anomalies in renewable energy systems using autoencoder-based detectors. For example, in [55]–[57], the authors have proposed an autoencoder neural network to analyze anomalies of wind turbine components using power generation or other SCADA data. However, the basic autoencoder-based anomaly detector is based on thresholding of residuals (reconstruction errors) using a Euclidean distance metric. This does not account for significant dependencies between measurements, such as the spatial and temporal correlations of renewable resources [58]. The mismatch between the detector design and the features of data could have a negative impact on detection sensitivity and localization performance.

In view of this, some authors have proposed using the Mahalanobis distance to measure *residuals* and thus acquire more accurate classification boundaries for autoencoder-based anomaly detectors [59]. Authors of [60]–[62] reported autoencoder-based wind turbine fault detectors using the Mahalanobis distance. However, it has not yet been investigated how the modified detection boundaries impact the anomaly localization performance.

Apart from adjusting residuals, the correlated renewable generation data, which are the *input* of autoencoder networks, can be decorrelated and standardized (i.e., whitened

[63]) before fed into the autoencoder. This technique has been considered in the field of computer vision, with a focus on image and video data sets, for example, image retrieval [64] and object recognition [65]. However, there has been little quantitative analysis of detection sensitivity and localization performance improvement in the context of utilizing an autoencoder-based detector with input data whitening [66]. Further, what is not yet clear is the impact of processing the inputs and residuals *together* on the capacity of a detector.

3.1.2. CONTRIBUTION AND OUTLINE

This chapter bridges these identified gaps by investigating the impact of whitening input data and residuals and quantifying the improvement of detection sensitivity and localization performance using our proposed metrics. This is done in the context of a high-dimensional renewable energy use case. The main contributions of this chapter are listed below:

1) Comparative studies of different data processing methods, neural network configuration schemes, and whitening matrix selections are carried out, and their influences on anomaly detection sensitivity and localization accuracy are quantified using a variety of metrics.

2) We propose a combined whitening of the input features *and* of autoencoder residuals, which is shown to maximize detection sensitivity in a case study on correlated high-dimensional renewable power generation.

3) A combination of input feature standardization and *ZCA-cor*- or *ZCA*-based residual whitening is shown to enhance the visibility of anomalies and thus achieve an outstanding localization performance of an anomaly detector. The performance is further enhanced by a tunable offset to the whitening transformation.

Section 3.2 first formulates the problems of detecting anomalies of dependent power system measurements and introduces different whitening matrices for different data processing steps. Then a data whitening scheme is proposed along with the anomaly localization metrics. Section 3.3 utilizes wind farm power generation as the case study, evaluating the influence of different data processing options, neural network configuration schemes, and whitening matrix selections on anomaly detection and localization performance of the autoencoder-based anomaly detector.

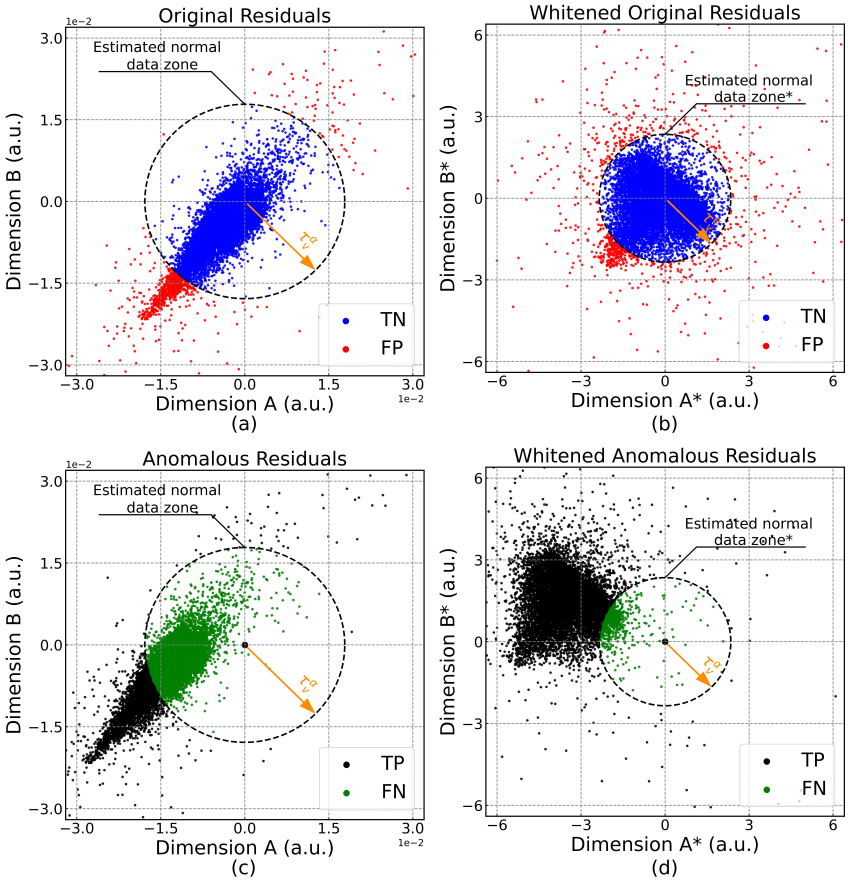

Figure 3.1: Illustrative two-dimensional distributions of (a) original residuals, (b) whitened original residuals, (c) synthetic anomalous residuals obtained by shifting, and (d) anomalous residuals whitened according to the normal data. The residuals are classified as TN (True Negatives), FP (False Positives), TP (True Positives), and FN (False Negatives) by comparing with a 95% threshold (α=95) calculated on a validation set.

3.2. DETECTOR ENHANCEMENTS

3.2.1. PROBLEM FORMULATION

Power system measurements exhibit spatial-temporal dependencies. For example, due to geographic factors, the scale and irradiance of renewable resources such as wind and solar are spatially dependent within a given region [58]. Autoencoder-based neural networks are trained to replicate these correlated inputs on the output side with minimal reconstruction errors.

Dependencies in inputs may also lead to correlated residuals. An example in Fig. 3.1 shows two dimensions of the residuals obtained in the case study of Section 3.3. The residuals of normal test data shown in Fig. 3.1 (a) are classified into true negatives (TN) and false positives (FP) by the threshold τ_v^α. However, the assumption of a circular 'estimated normal data zone' is not appropriate for this ellipsoidal distribution. Fig. 3.1 (c) illustrates this with simulated anomalous data that is obtained by shifting the residuals. Many clearly anomalous points are within the normal circle and therefore not detected (False Negatives, FN). This reduces the probability that an actual anomaly is identified: the true positive rate (TPR), also known as detection sensitivity. The illustration in two dimensions also applies to residuals in higher dimensions.

3.2.2. DATA WHITENING SCHEMES

In view of the elliptically distributed residuals and concomitant errors in anomaly detection, whitening (also known as sphering) the observations is a promising approach to improve detection performance. By removing the correlations between the residual components, the n-ball may better describe the normal data distribution, and anomalies may be detected more accurately. The potential effectiveness of this approach is depicted in Fig. 3.1 (b,d). Whitening can be applied in three different combinations of two approaches as:

- Whitening of the input data;
- Whitening of generated residuals;
- Combined whitening of the input data and residuals.

We first summarize the properties of the whitening transformation. Consider a random vector $Z = (z_1, \ldots, z_n)^T$, with the (non-singular) covariance matrix $\mathrm{Cov}(Z, Z) = \Sigma \in \mathbb{R}^{n \times n}$. We define the (also non-singular) *whitening transformation matrix* $W \in \mathbb{R}^{n \times n}$ such that

$$V = (V_1, \ldots, V_n)^T = WZ, \tag{3.1}$$

where the elements of the random vector V are uncorrelated and have unit variance: $\mathrm{Cov}(V, V) = \mathbf{1}$. We determine constraints on W by expanding

$$\begin{aligned} \mathrm{Cov}(V, V) &= \mathrm{E}[WZ(WZ)^T] - \mathrm{E}[WZ]\mathrm{E}[(WZ)^T] \\ &= W\big(\mathrm{E}[ZZ^T] - \mathrm{E}[Z]\mathrm{E}[Z^T]\big)W^T = W\Sigma W^T. \end{aligned} \tag{3.2}$$

This implies the constraint $W\Sigma W^T = \mathbf{1}$. As W is invertible, we multiply with W^{-1} and $(W^T)^{-1}$ from the left and right, respectively, and find after inversion:

$$W^T W = \Sigma^{-1}. \tag{3.3}$$

This does not determine the whitening matrix W uniquely. Among the infinite possible options of whitening matrices W, a few are commonly used [67]. In this chapter, four approaches are studied: *PCA*, *ZCA*, *Cholesky*, and *ZCA-cor* [63].

The *PCA whitening* transformation is a widely used sphering approach due to its close relation to *principle component analysis* (*PCA*)[68]. It can be regarded as rescaling variances of all dimensions to one after a *PCA* procedure that omits the customary dimension reduction. The whitening matrix is

$$W_{PCA} = \Lambda^{-1/2} U^T, \tag{3.4}$$

where $\Lambda \in \mathbb{R}^{n\times n}$ is a diagonal matrix with the eigenvalues of covariance matrix Σ and the columns of $U \in \mathbb{R}^{n\times n}$ are the corresponding eigenvectors. It is closely related to the *ZCA* approach, which uses U to transfer the *PCA* whitened data back to the original coordinate system[63]:

$$W_{ZCA} = U\Lambda^{-1/2} U^T. \tag{3.5}$$

The *Cholesky whitening* transformation is defined as

$$W_{Chol} = L^T, \tag{3.6}$$

where $L \in \mathbb{R}^{n\times n}$ is a lower triangular matrix with positive diagonal entries, obtained by *Cholesky* decomposition of the precision matrix (inverse covariance matrix): $\Sigma^{-1} = LL^T$. The final sphering approach considered in this chapter is *ZCA-cor whitening* transformation[63]. It uses the whitening matrix

$$W_{ZCA-cor} = S^{-1/2} V^{-1/2}, \tag{3.7}$$

where $V \in \mathbb{R}^{n\times n}$ is the diagonal variance matrix and $S \in \mathbb{R}^{n\times n}$ denotes the correlation matrix (so that $\Sigma = V^{1/2} S V^{1/2}$). The *ZCA-cor whitening* approach maximizes the correlation of whitened and original components [63]. Unlike W_{ZCA}, $W_{ZCA-cor}$ is in general asymmetric.

In this chapter, we consider both the detection and localization performance of the anomaly detector. As the whitening procedure is transparent to the calculation of residual vectors (the squared vector is used), the particular choice of W mostly affects the

localization performance. It may also affect the detection performance, albeit indirectly, if whitening is applied to the input data, thus affecting the training of the autoencoder.

INPUT WHITENING

The whitening transformation can be utilized to remove correlations from the data used for training and testing. First, this enables the autoencoder network to learn from less redundant inputs, which is generally desirable[65]. But more importantly, we hypothesize that the reduction in the input correlation may propagate to the residuals.

We consider a *whitened input data point*

$$z_w = W_z(z - \mu_z), \tag{3.8}$$

with W_z, the z-space whitening matrix, computed from the sample covariance of the training data z. The data is also (optionally) centered on the data mean μ_z. The residual $r_w \in \mathbb{R}^n$ is defined as $r_w = z_w - \hat{z}_w$, where $\hat{z}_w$ is the reconstructed data point. Inverting the whitening procedure gives $\hat{z}_a = W_z^{-1}\hat{z}_w + \mu_z$, which may be compared with the original z. The reconstruction error of the whitened data $\|r_w\|_2$ can be related to that of $r_a = z - \hat{z}_a$ as:

$$\begin{aligned} \|r_w\|_2 = \|z_w - \hat{z}_w\|_2 &= [(z - \hat{z}_a)^T W_z^T W_z (z - \hat{z}_a)]^{1/2} \\ &= [(z - \hat{z}_a)^T \Sigma_z^{-1} (z - \hat{z}_a)]^{1/2} \triangleq \|r_a\|_{\Sigma_z^{-1}}. \end{aligned} \tag{3.9}$$

Compared with (2.7), by taking correlations of original inputs into account, we are effectively measuring the Mahalanobis distance [69] between z and $\hat{z}_a$ instead of their standard Euclidean length. Whitening of the input data thus affects both the representation of the training data as well as the loss function used during training.

RESIDUAL WHITENING

In contrast with applying whitening transformation before feeding data into the neural network, *residual whitening* reshapes the distribution of residuals *for a given trained autoencoder*. Concretely, the raw residual $r = [r_1, \ldots, r_n]^T$ is whitened as

$$r_s = W_r(r - \mu_r). \tag{3.10}$$

Here, the whitening matrix $W_r \in \mathbb{R}^{n \times n}$ is computed on the sample covariance of residuals from the *validation* data, because the training data set is used to train the autoencoder itself. μ_r represents the mean of raw residuals in the validation set, which should be

approximately zero if the RMSE loss function was used during training. Accordingly, the reconstruction error is given by

$$\|r_s\|_2 = [(r-\mu_r)^T \Sigma_r^{-1}(r-\mu_r)]^{1/2} \triangleq \|r-\mu_r\|_{\Sigma_r^{-1}}. \tag{3.11}$$

Slightly different from (3.9), the reconstruction error in (3.11) denotes the Mahalanobis distance of a residual r from a set of residuals with mean μ_r and covariance matrix Σ_r.

COMBINED DATA-RESIDUAL WHITENING

Apart from the above two intuitive manners, we propose to utilize the data and residual whitening schemes together as a combined whitening procedure. Specifically, it further reshapes the residuals generated from the reconstructed whitened inputs. We define $r_t \in \mathbb{R}^{n\times 1}$ to represent the residuals after implementing combined whitening transformation. The corresponding reconstruction error is denoted as

$$\|r_t\|_2 = \|W_{r_w}(r_w-\mu_{r_w})\|_2 \triangleq \|W_z r-\mu_{r_w}\|_{\Sigma_{r_w}^{-1}}, \tag{3.12}$$

where $\mu_{r_w} \in \mathbb{R}^{n\times 1}$ and $\Sigma_{r_w} \in \mathbb{R}^{n\times n}$ refers to the mean value and covariance matrix of r_w, respectively, and $W_{r_w} \in \mathbb{R}^{n\times 1}$ stands for its whitening matrix. In this way, it doesn't only force the autoencoder to learn important features with less redundant training data but also spheres the residuals in case of remaining residual correlations.

3.2.3. ANOMALY LOCALIZATION METRICS

In many scenarios, when a likely anomaly has been detected, it is important to also identify which observation(s) triggered the anomaly detector. They may indicate a component malfunction or the source of the physical disturbance. In the case study that follows, we will show that with a well-chosen whitening procedure, the values of the residual vector can be used to pinpoint the anomaly: the highest absolute residuals are the most likely locations of anomalies.

To quantify the dependability of the localization performance, we propose three metrics. The first of these is the *RMS Ratio*, which denotes the ratio of root mean square value of anomalous dimensions to that of normal dimensions in residual vectors. This is given by

$$\textit{RMS Ratio} = \frac{1}{m}\sum_{l=1}^{m}\{\Big[\frac{1}{|\mathscr{A}|}\sum_{j\in\mathscr{A}}(r_j^{(l)})^2\Big]^{\frac{1}{2}} / \Big[\frac{1}{|\mathscr{N}|}\sum_{j\in\mathscr{N}}(r_j^{(l)})^2\Big]^{\frac{1}{2}}\}, \tag{3.13}$$

where $r_j^{(l)}$ denotes the j^{th} element of the residual of the l-th test data point. $\mathscr{A}$ is the set of anomalous dimensions (e.g., malfunctioning devices), and $\mathscr{N}$ is the set of non-anomalous dimensions. Moreover, m refers to the total number of records in the test set. The *RMS Ratio* can be used to estimate if the anomalous stand out on average. A second, more stringent metric is introduced as well: *Gap Ratio*, which measures the average ratio of the smallest anomalous dimension to the largest normal dimension (absolute value). It is defined as

$$Gap\ Ratio = \frac{1}{m}\sum_{l=1}^{m}\{\min_{j\in\mathscr{A}}|r_j^{(l)}|/\max_{j\in\mathscr{N}}|r_j^{(l)}|\}. \tag{3.14}$$

The third metric, ordinal consistency rate (*OCR*), calculates the proportion of samples for which the smallest absolute value of an *anomalous* dimension exceeds the largest absolute value of a *non-anomalous* dimension:

$$OCR = \frac{1}{m}\sum_{l=1}^{m}\mathbb{1}_{\min_{j\in\mathscr{A}}|r_j^{(l)}|>\max_{j\in\mathscr{N}}|r_j^{(l)}|}. \tag{3.15}$$

3.2.4. DESIGN OF THE ANOMALY DETECTOR

We integrate the anomaly detection mechanism and whitening schemes to give data processing options as well as explain which data is used in different stages. The proposed data flow and its transformation processes in the autoencoder neural network-based anomaly detector are depicted in Fig. 3.2 with the following five steps.

DATA PARTITION

Given the historical data set X, the first step is to divide observations into training, validation, and test data set as X_t, X_v, and X_e with a specific ratio.

INPUT PRE-PROCESSING

In this step, statistics of the training data (mean, range, covariance) are computed, and these values are used for input processing of training, validation, and test data according to the selected method, e.g. input whitening shown in (10).

TRAINING AND RECONSTRUCTION

The weight matrices K and bias vectors b are updated iteratively to minimize the reconstruction loss in (2.7). Afterwards, the trained autoencoder neural network is utilized to reconstruct the validation and test set to $\hat{X}'_v$ and $\hat{X}'_e$. Then, the corresponding residual sets r'_v and r'_e are calculated, such as r_w in (11).

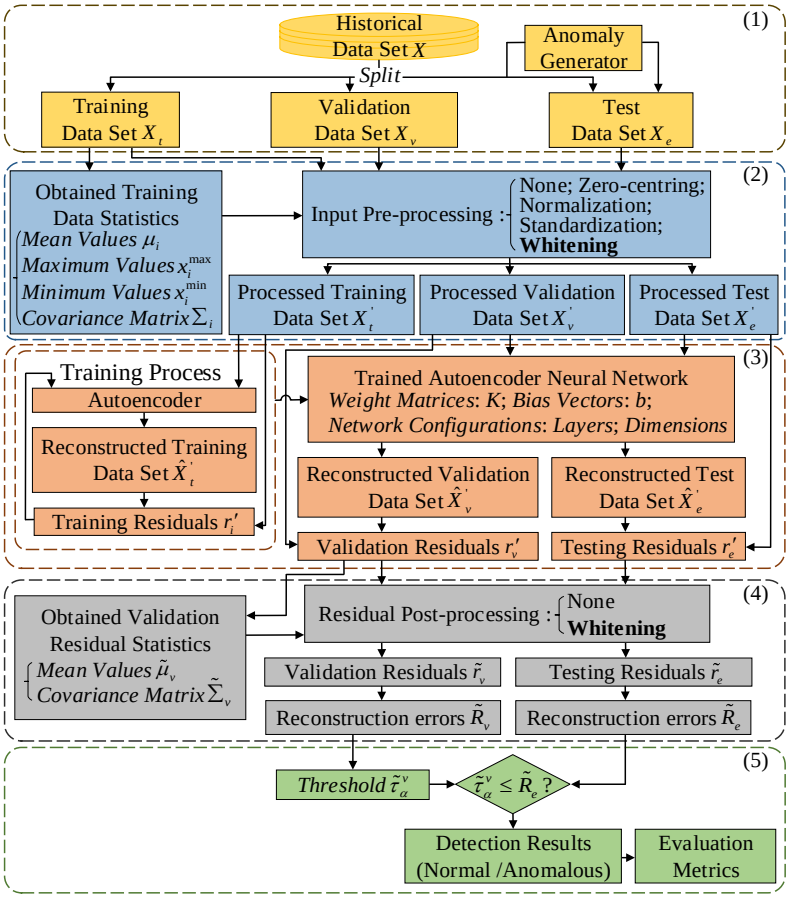

Figure 3.2: The proposed framework of data flow in the autoencoder neural network-based anomaly detector.

RESIDUAL POST-PROCESSING

If residual whitening is employed, the validation residuals are used to compute whitening transformations. After executing the residual whitening transformation shown in (3.10), the reconstruction errors of the validation set $\tilde{R}_v$ and test set $\tilde{R}_e$ are calculated as (3.11).

DETECTION PERFORMANCE EVALUATION

The anomaly threshold $\tilde{\tau}_v^\alpha$ is obtained as a quantile of the reconstruction errors $\tilde{R}_v$, corresponding to the desired true negative rate $\alpha\%$. The test data are classified by comparing reconstructions $\tilde{R}_e$ with the threshold $\tilde{\tau}_v^\alpha$. Consequently, the performance of the anomaly detector is assessed by calculating the evaluation metrics based on the pre-

dicted normal, anomalous states, and actual states.

3.3. DETECTION OF ANOMALOUS WIND FARM GENERATIONS

In this section, the impacts of different data processing options, neural network configurations, and whitening matrix selections on anomaly detection and localization capacity of an autoencoder-based detector are investigated. To do so, we conduct a case study on power generation from distributed wind farms. Although introduced anomalies are synthetic, the wind farm output is based on reanalysis data of historical wind speeds. This use of maximally realistic high-dimensional data ensures that the data won't be easily compressed into a low-dimensional manifold by an autoencoder.

Specifically, we first describe the process of modeling normal data patterns and then formulated anomalous scenarios. For these scenarios, we make anomaly detection performance comparisons by implementing various combinations of data transformations including whitening. Next, anomaly locating capacity evaluation is conducted by comparing four whitening transformations: *PCA*, *Cholesky*, *ZCA*, and *ZCA-cor*.

3.3.1. EXPERIMENT SCENARIO FORMULATION

Renewable power generation from spatially distributed wind farms is an increasingly relevant source of energy. The power output of each wind farm is highly variable due to variations in wind speed, but this may obscure other factors causing reduced performance. Given this, experiments were conducted to test if our proposed mechanism can detect and localize anomalies in the power output of wind farms with satisfactory capacity. Notably, without knowing any model-related information about wind farms and relying on the neural network-based data-driven methodology only, our proposed anomaly detection mechanism was trained on historical operation data and tested on both normal and anomalous scenarios. In this chapter, we generated anomalous scenarios as reductions in the power output of one or more wind farms. These could reflect unexpected malfunctions, disturbances, unscheduled outages, or unreported maintenance activities (from the perspective of system operators).

A realistic wind power data set was constructed as follows. We virtually placed a 100MW wind farm at each center of the 99 municipalities located in the North and South Holland provinces of the Netherlands [70], [71]. The wind power output was simulated on the basis of historical wind speeds at the 99 locations, obtained by MERRA-2 reanal-

Table 3.1: Data Processing Method Combinations.

	Input Pre-processing	Residual Post-processing
$\boldsymbol{P_{NN}}$	None	None
$\boldsymbol{P_{SN}}$	Standardization	None
$\boldsymbol{P_{WN}}$	Whitening	None
$\boldsymbol{P_{NW}}$	None	Whitening
$\boldsymbol{P_{SW}}$	Standardization	Whitening
$\boldsymbol{P_{WW}}$	Whitening	Whitening

ysis and available from renewables.ninja [72]. After obtaining the historical wind data [73], the associated generated power outputs were calculated as described in[74]. Ultimately, the whole generated data set $X \in \mathbb{R}^{99\times 87648}$, which includes 10 years' (2009-2018) hourly outputs of 99 wind farms, was divided into the training set X_t, validation set X_v and test set X_e in blocks of one week with the proportion of 6, 2, and 2 years.

The autoencoder encoded and decoded the 99-dimensional data from the input to the output layer. Both the encoder and decoder had 3 hidden layers with 200 neurons, connected to a bottleneck layer with variable size. The bottleneck size is indicated as B_n, where n is the number of neurons in the bottleneck layer. The ReLU activation function was used, and the Adam Optimizer [51] was utilized with default settings to iteratively optimize the value of weight matrices K and bias vectors b. In this research, 5×10^3 training epochs were used. The batch size and learning rate for training were 64 and 5×10^{-5}, respectively. Training and testing of the autoencoder were conducted using `tensorflow`.

For a comparative study of anomaly detection and localization performance, in the following subsections, we made use of different combinations of pre-processing methods for inputs and post-processing schemes for residuals. The combinations, denoted as P_{xy}, are listed in Table 3.1.

3.3.2. IMPACT OF WHITENING TRANSFORMATION

Fig. 3.3 depicts the correlation coefficients of the testing residuals $\tilde{r}_e$ for 9 out of the 99 dimensions, for a variety of data processing methods. When the input data is standardized,

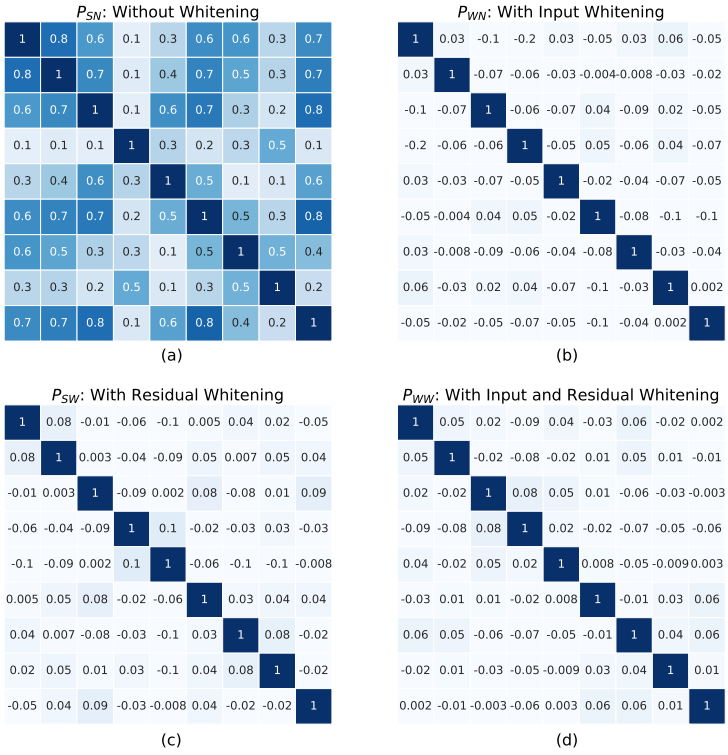

Figure 3.3: The correlation coefficient matrices of normal testing residuals $\tilde{r}_e$ when utilizing four data processing combinations. The dimensions shown correspond to nine virtual wind farms (B_{64} is utilized).

but no whitening is performed (P_{SN}), high correlations among different dimensions are visible, which implies 'elliptically' distributed residuals according to the analysis in Section 3.2.1.

As expected, whitening effectively reduces these correlations. Whitening of the input data (P_{WN}) drastically reduces correlations between the residuals. Correlations are slightly lower still when whitening is applied directly to the residuals (P_{SW}). Note that the pairwise correlations are not zero due to differences between the validation set (used to determine the post-whitening matrix) and the test set. Finally, applying whitening on the inputs and the outputs (P_{WW}) produces similarly small correlations.

3.3.3. ANOMALY DETECTION PERFORMANCE EVALUATION

Both the processing methods (P_{xy}) and the autoencoder configuration (B_x) influence the sensitivity of anomaly detection. This impact will be quantified in this section. For

these tests, anomalies were generated by modifying the test data as follows. For each data point, we randomly selected one wind farm out of 99 and reduced its power output by a given amount (the *anomaly magnitude*, abbrev. *am*). This approach yields an anomalous test set consisting of 17,544 non-anomalous data points and an equal number of anomalous data points.

RECEIVER OPERATING CHARACTERISTIC CURVES

Receiver Operating Characteristics (ROC) and the corresponding Area Under the Curve (AUC) were used to quantify the sensitivity and specificity of anomaly detection, as a function of the autoencoder structure, data processing method, and anomaly magnitude. ROC curves were constructed by varying the threshold τ_v^α. For all cases, *ZCA* was selected as the whitening matrix, and an anomaly magnitude of 10% was used.

For the first experiment, we used default processing P_{SN}: standardizing the input data and detecting anomalies from unprocessed residuals. Comparing the performance of all layer dimension configurations B_x, we can observe in Fig. 3.4 (a) that autoencoder networks configured as B_{32}, B_{64} and B_{96} have better detection performance. Specifically, at each false positive rate, the true positive rate (sensitivity) of these detectors are higher than others'. Accordingly, they also have larger AUC. This indicates that optimal detection is achieved with fairly wide autoencoders. The configuration B_{64} was used for all the following experiments.

A comparison of the anomaly detection performance of different data processing approaches P_{xy} is shown in Fig. 3.4(b). We can observe that, i) detectors equipped with whitening transformation (P_{WN}, P_{NW}, P_{SW}, and P_{WW}) have higher anomaly detection sensitivity than the others; ii) performing whitening only on the inputs (P_{WN} and P_{WW}) renders higher detection sensitivity than whitening approaches performed to the residuals (P_{NW} and P_{SW}); iii) the detector using combined whitening (P_{WW}) slightly outperforms the detector just utilizing input whitening (P_{WN}). It can be concluded that the combined whitening approach P_{WW} is the best choice to improve overall anomaly detection sensitivity.

Moreover, we investigated the ability to detect anomalies of various magnitudes, using the selected processing strategy P_{WW}. Fig. 3.4 (c) shows that, as the anomaly rate increases from 1% to 30%, the ROC curves and AUC improve, reaching very high levels from 10%.

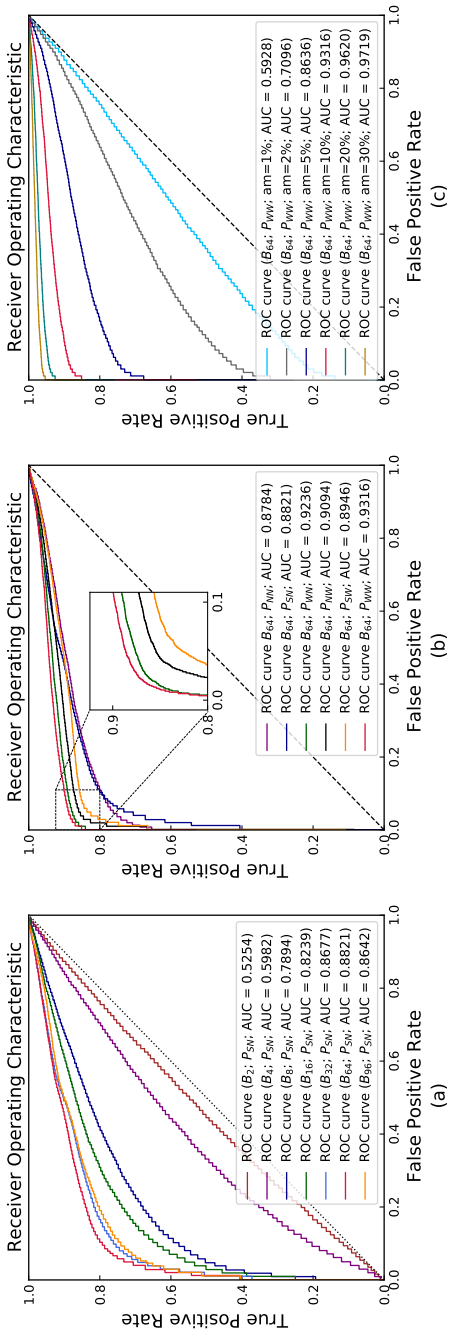

Figure 3.4: Test set receiver operating characteristics, using different configurations (B_x), data processing methods (P_{xy}), and anomaly magnitudes.

Table 3.2: Detection performance comparison by multiple metrics (α = 99, anomaly magnitude = 10% and B_{64} is utilized).

Processing method	PPV	TNR	ACC	F_1-score
$\boldsymbol{P_{NN}}$	98.07%	98.89%	82.30%	78.70%
$\boldsymbol{P_{SN}}$	97.06%	98.77%	69.68%	57.23%
$\boldsymbol{P_{WN}}$	**98.80%**	**98.98%**	91.52%	90.83%
$\boldsymbol{P_{NW}}$	98.44%	98.86%	82.82%	79.50%
$\boldsymbol{P_{SW}}$	98.22%	98.80%	82.66%	79.32%
$\boldsymbol{P_{WW}}$	98.70%	98.88%	**92.01%**	**91.42%**

DETECTION PERFORMANCE EVALUATION BY MULTIPLE METRICS

In addition to the detection sensitivity (true positive rate), we evaluated the performance of the anomaly detector by multiple metrics, namely precision (PPV), specificity (TNR), accuracy (ACC), and F_1-score. Interested readers can refer to [75] for a detailed introduction to the metrics. For all cases, we used an anomaly magnitude of 10%, $\alpha = 99$, and layer configuration B_{64}. The experimental results are shown in Table 3.2. In all cases, the TNR is close to 99%, by the choice of the threshold. The PPV scores are high across all processing methods, but a closer look at the ACC and F_1 metrics show that - perhaps surprisingly - the P_{SN} processing scheme is least dependable, by a large margin. The schemes using whitening of input data outperform all others, with a slight edge for the combined whitening procedure.

To investigate stability of the stochastic training process, model training was performed 15 times, and the variability of the sensitivity (true positive rate) was monitored. Whitening at the pre-processing stage resulted in higher sensitivity *and* a narrower range of results. The result is shown in Fig. 3.5

3.3.4. ANOMALY LOCALIZATION PERFORMANCE EVALUATION

In addition to quantifying the ability to detect anomalies (a binary classification), we next investigated the ability to localize the source of an anomaly, and how this depends on the configuration of the detector.

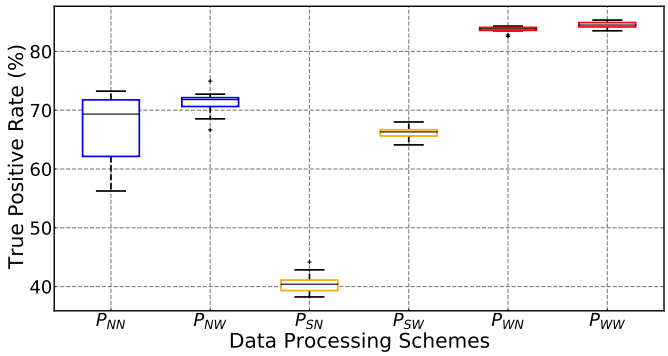

Figure 3.5: Distribution of true positive rates across 15 training runs (α = 99, anomaly magnitude = 10%, and B_{64} is utilized).

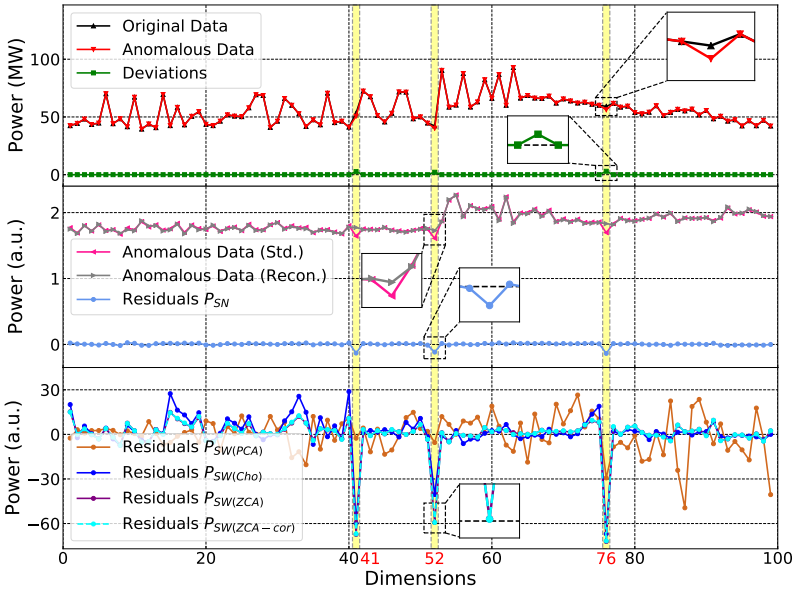

Figure 3.6: Illustrative example that depicts the effect of different whitening approaches on anomaly locating capacity (anomaly magnitude = 5%; P_{SN}, P_{SW}, and B_{64} are utilized).

VISUAL COMPARISON OF LOCALIZATION PERFORMANCE

We first considered an illustrative example of anomaly localization, in which the output of three wind farms (numbers 41, 52, and 76) was reduced by 5%. The input data is shown in Fig. 3.6 (top panel), where the anomalous locations are indicated in yellow.

Looking at the residuals of standardized data (P_{SN}, middle panel), it can be seen that

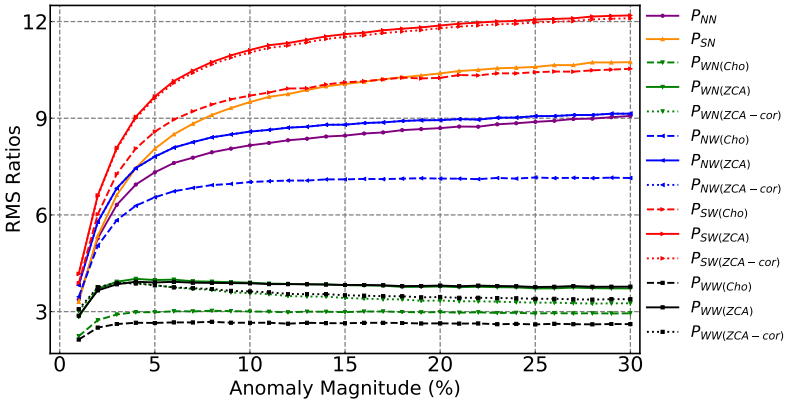

Figure 3.7: Root mean square (RMS) ratios of the anomalous to normal dimensions for data processing/whitening combinations (B_{64} is utilized).

the residuals of the three anomalous dimensions stand out. In the bottom panel, we can observe that the four whitening approaches affect the residual signals in different ways. *PCA* whitening mixes all coordinates and fully obscures the connection between the original perturbations and residuals. As a result, it will not be considered in the follow-up analysis. In contrast, the *Cholesky*, *ZCA*, and *ZCA-cor* whitened residuals all have peaks that are consistent with the actual anomalous dimensions, but the *Cholesky* method also produces residuals in non-anomalous locations (e.g., a peak at dimension number 40).

STATISTICAL LOCALIZATION PERFORMANCE COMPARISON

For each point in the test set, we randomly selected 3 out of 99 wind farms and applied power reductions to generate anomalous test vectors, resulting in test sets of 17544 data points for each anomaly rate. Fig. 3.7 depicts the *RMS Ratio* for various data processing schemes, as a function of anomaly magnitude, and Table 3.3 shows numerical results for all three anomaly metrics for a fixed anomaly magnitude of 5%.

The most striking observation is that methods that perform whitening at the pre-processing stage (P_{Wx}) scored significantly worse on all localization metrics. Apparently, the mixing of features before encoding helps to improve detection sensitivity (previous section), but is detrimental to localization performance. Moreover, for any data processing combination, both *ZCA* processing schemes scored higher than the *Cholesky* whitening scheme, and the best scores were obtained when the *ZCA* and *ZCA-cor* schemes are

Table 3.3: Localization capacities of using different data processing methods (anomaly magnitude = 5% and B_{64} is utilized).

Processing method	Whitening matrix	RMS Ratio	Gap Ratio	OCR
$\boldsymbol{P_{NN}}$	/	7.32	1.60	73.18%
$\boldsymbol{P_{SN}}$	/	8.05	1.97	78.48%
$\boldsymbol{P_{WN}}$	*Cholesky*	2.98	0.37	7.17%
$\boldsymbol{P_{WN}}$	*ZCA*	3.98	0.68	23.40%
$\boldsymbol{P_{WN}}$	*ZCA-cor*	3.82	0.61	18.22%
$\boldsymbol{P_{NW}}$	*Cholesky*	6.55	1.39	68.04%
$\boldsymbol{P_{NW}}$	*ZCA*	7.81	1.65	75.98%
$\boldsymbol{P_{NW}}$	*ZCA-cor*	7.81	1.65	75.74%
$\boldsymbol{P_{SW}}$	*Cholesky*	8.58	1.59	71.78%
$\boldsymbol{P_{SW}}$	*ZCA*	**9.68**	**1.98**	82.23%
$\boldsymbol{P_{SW}}$	*ZCA-cor*	9.62	1.97	**82.60%**
$\boldsymbol{P_{WW}}$	*Cholesky*	2.65	0.30	2.44%
$\boldsymbol{P_{WW}}$	*ZCA*	3.90	0.65	20.52%
$\boldsymbol{P_{WW}}$	*ZCA-cor*	3.82	0.61	18.27%
$\boldsymbol{P_{SW}}$	*ZCA* - c	**9.72**	**2.07**	84.65%
$\boldsymbol{P_{SW}}$	*ZCA-cor* - c	9.67	2.06	**85.08%**

used for post-processing, in combination with standardization for pre-processing (P_{SW}). Here, *ZCA* scored very slightly higher on the *RMS ratio* and *gap ratio* metrics (typical separation), and *ZCA-cor* attained the highest scores on the *OCR* metric (ordering).

Finally, an additional enhancement of the method was introduced. The residual whitening transformation (3.10) mixes signals between dimensions. This may cause a peak (positive or negative) in one or more dimensions to affect the average value of other dimensions. In order to better separate this signal from the background, we applied a constant offset to the whitening matrix (3.10) as follows:

$$r_{s(c)} = (W_r - c\mathbb{1})r. \tag{3.16}$$

Here, $\mathbb{1}$ is a matrix of ones and c is a constant to be defined, so that a multiple of the sum-of-residual values ($\sum_i r_i$) is subtracted from the whitened feature vector. The change of localization performance as a function of c is shown in Fig. 3.8. An overall improvement is obtained for values larger than zero, although *RMS ratio* decreases after an initial increase. Results for the value $c = 5$ are included in Table 3.3.

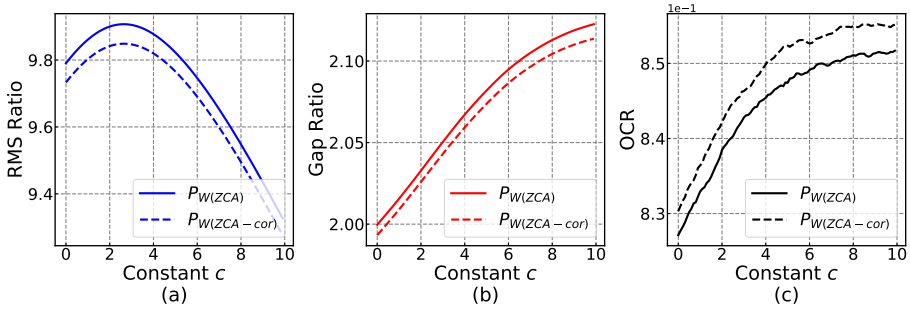

Figure 3.8: Localization performance as a function of whitening offset c.

4

Data Generation Using a Conditional Variational Autoencoder

For planning of power systems and for the calibration of operational tools, it is essential to analyze system performance in a large range of representative scenarios. When the available historical data is limited, generative models are a promising solution. In chapter 2 and 3, we have investigated the performance of autoencoder-based anomaly detectors. Similar to the task of anomaly detection, data generation is also an application based on acquiring the features of historical data. However, modeling both marginal distribu-

This chapter is based on the following published work:

[76] C. Wang, E. Sharifnia, Z. Gao, S. H. Tindemans, and P. Palensky, "Generating multivariate load states using a conditional variational autoencoder", *presented* in *XXII Power Systems Computation Conference (PSCC 2022)*, Porto, Portugal, 2022 and *published* in *Electric Power Systems Research*, vol. 213, p. 108603, 2022.

[77] C. Wang, S. H. Tindemans, and P. Palensky, "Generating contextual load profiles using a conditional variational autoencoder", in *2022 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, Novi Sad, Serbia, 2022, pp. 1–6. DOI: 10.1109/ISGT-Europe54678.2022.9960309

The cGAN model shown in Section 4.3 for a comparative study is developed by Zhi Gao.

Section 4.4 is based on results from Ensieh Sharifnia.

tions and multivariate dependencies of the historical data is challenging. In this chapter, a multivariate load state generating model on the basis of a conditional variational autoencoder (CVAE) neural network is proposed. This neural network is a variant of the autoencoder described in the previous two chapters. Going beyond common CVAE implementations, the model includes a stochastic variation of output samples under given latent vectors and co-optimizes the parameters for this output variability. It is shown that this improves the statistical properties of the generated data. The quality of generated multivariate loads is evaluated using univariate and multivariate performance metrics. A generation adequacy case study on the European network is used to illustrate the model's ability to generate realistic tail distributions. The experiments demonstrate that the proposed generator outperforms other data generating mechanisms. Moreover, in this chapter, we test the generative model for load profiles of industrial and commercial customers, which is challenging due to the highly variable nature of such profiles. The experimental results demonstrate the CVAE model can capture temporal features of historical load profiles and generate 'realistic' data with satisfying univariate distributions and multivariate dependencies.

4.1. INTRODUCTION

4.1.1. STATE OF THE ART OF DATA GENERATORS

In order to plan power systems and calibrate operational tools, it is essential to analyze system performance across a large range of representative scenarios [19], [20]. Historical data is a key source of such scenarios, but when the available data set is too small for the desired application or when it cannot be made available for privacy reasons, it becomes valuable to have a model that can generate relevant data in abundant quantities. The challenge is that generated scenarios should embody both univariate distributions and multivariate inter-dependencies of the historical data [78]. Compared to the work in Chapter 2 and Chapter 3, the generative model considers the joint probability of a given observation, while anomaly detection is essentially a discriminative task, i.e., estimating the conditional probability that a given observation is normal or abnormal. However, both of them are based on capturing the features of observable variables. In this research, efforts are paid to generate power system load data.

A common approach has been to fit parametric probabilistic models to historical scenarios. In [79], *Gaussian mixture models* (GMM) have been proposed to augment load data in distribution networks. More recently, a load generator has been designed using time-varying queuing models [80]. Due to the curse of dimensionality, it is especially challenging to use parametric methods for the generation of high-dimensional states [81, chapter 3]. *Copula*-based models are one class of generative models that do scale to higher dimensions, either using the *Gaussian copula*, or by 'stacking' copulas in a vine structure, possibly in combination with a dimension-reduction scheme [78]. As vine-based copula models are highly asymmetric and therefore prone to bias, it is appealing to investigate 'native' high-dimensional models, such as neural networks.

The *generative adversarial network* (GAN) [32] and *variational autoencoder* (VAE) [82] are two representative neural network-based generative models. The generator and discriminator of GAN are trained in the form of contesting with each other. After training, new data are generated with similar statistics as the training set. GAN-based load data generators have been provided in [83], [84]. The *variational autoencoder* (VAE) [82] is an unsupervised machine learning model with an extra constraint for the lower-dimensional latent space codes, which can be considered a variant of the autoencoder. The introduction of unobserved auxiliary latent space codes is motivated by the thoughts that 'significant' information tends to reside on a lower-dimensional manifold, and the

known relations between latent space code and the observable high-dimensional data can be used to simplify the generative model.

The VAE model has been successfully used in generating electricity load series, such as theft detection [22] and electric vehicle load profiles [23]. However, the validation of generated data often remains limited to visual comparisons, which is not straightforward for snapshots of larger and more complex electricity systems. Moreover, most VAE implementations do not make full use [85] of the flexibility permitted by the mathematical framework in [82]. Output noise tuning and training [86], [87] has only recently been considered, with a focus on image and video data sets. In power system related data generation applications, the output noise parameter is usually treated as a hyperparameter (i.e. a preset value that controls the learning process) [88] and noise is not actually inserted into samples [22], [23], [89]–[91].

4.1.2. CONTRIBUTION AND OUTLINE

This chapter bridges those identified gaps by investigating the impact of output noise and its parameterization, and by analyzing generated data using performance metrics. This is done for the VAE and the *conditional* VAE (CVAE), in the context of large-scale spatial load patterns of European countries. This lays the basis for the application of synthesizing load generation at lower aggregation levels, where consumption patterns are inherently more variable.

The main contributions of this chapter are:

1) We show how a sample-dependent output noise parameter can be co-optimized in the training process and how this noise is used in the generative process.

2) We put forward a set of data quality metrics for generative models, consisting of three statistical tests for univariate distributions and multivariate dependencies.

3) We introduce a simple multi-area adequacy assessment model that is used to test tail distributions.

4) Through comprehensive experiments, we show the performance and practicality of VAE- and CVAE-based load generators in comparison with *Gaussian copula* and *conditional generative adversarial network* (cGAN) models.

5) For an application to generate load profiles of individual electricity users, we evaluate the performance of the CVAE model under different combinations of time of

year and power exchange intensity conditions using an anonymized data set of 5,000 industrial and commercial customers.

Section 4.2 introduces the structure of the conditional variational autoencoder-based generative model and describes the training and generation processes of the model. The mathematical variations of different usage of output noise are elaborated. Section 4.3 investigates the impact of output noise, weighting factors, and generative models on the performance of generating European country-level load data. The distribution and correlation of the generated data are evaluated. Section 4.4 proposes a multi-area adequacy assessment to show the practicability of using the CVAE-based load states generator for assessing the potential system risk. Section 4.5 investigates the capacity of the CVAE-based generative model to generate load profiles of individual users, of which the loads are more variable.

4.2. DATA GENERATION MECHANISM

To generate load states similar to the univariate and multivariate interdependencies of historical data, a representative multivariate load state generation mechanism is proposed, which is based on the CVAE. The four options of whether the output noise parameters are co-optimized during the training process and added in the generation process are introduced.

4.2.1. CVAE-BASED GENERATIVE MODEL

The CVAE is a neural network architecture that is trained to learn the salient features of historical data by mapping (*encoding*) historical system states onto a lower-dimensional latent space where the latent distribution is approximately normal - and transforming latent vectors back (*decoding*) into a high-dimensional state space [92]. The decoder is used in conjunction with contextual information c to generate representative states (which can be omitted to obtain a regular VAE model). Consequently, the model is able to generate samples with a similar distribution to the historical data, by transforming normally distributed samples in the latent space back to the data space. We note that the latent (i.e. hidden) representation of a data point is used solely to facilitate reconstruction and synthesis. It does not need to be imbued with a particular meaning.

During training, the structure of the CVAE algorithm is depicted in Fig. 4.1a. The *encoder* maps the d-dimensional input data x to the code z in the lower-dimensional

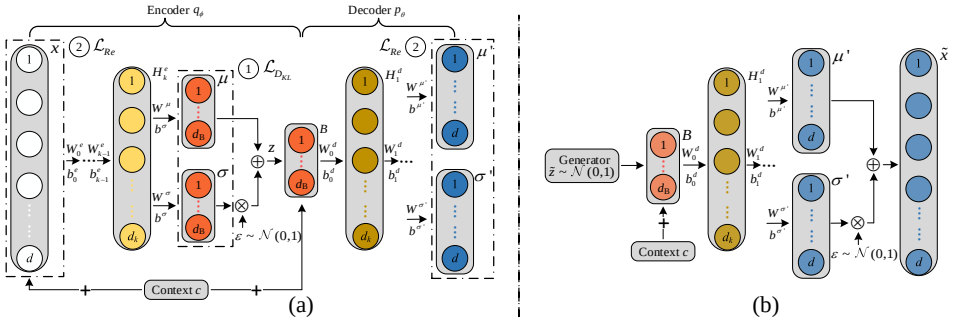

Figure 4.1: Schematic of the CVAE.

4

latent space through k hidden layers H_l^e, $l = 1,\ldots,k$. Weight matrices W_l^e, bias vectors b_l^e and the context c are utilized in the encoding process as[1]

$$\begin{pmatrix}\mu\\ \sigma\end{pmatrix} = \begin{pmatrix}W^\mu\\ W^\sigma\end{pmatrix}(a(W_k^e(\ldots a(W_1^e(x,c)+b_1^e)\ldots)+b_k^e)) + \begin{pmatrix}b^\mu\\ b^\sigma\end{pmatrix}, \tag{4.1a}$$

$$z = \mu + \epsilon \odot \sigma\,, \tag{4.1b}$$

where a represents an element-wise nonlinear activation function. Vectors μ and σ parameterize an input-dependent normal distribution in the latent space. The output z is sampled accordingly, using ϵ, a vector that is sampled from a standard normal distribution, and the Hadamard product $\odot$.

Mirroring the encoder network, the decoder maps the sampled latent space code z to the d-dimensional output data $\hat{x}$ using

$$\begin{pmatrix}\mu'\\ \sigma'\end{pmatrix} = \begin{pmatrix}W^{\mu'}\\ W^{\sigma'}\end{pmatrix}(\ldots a(W_1^d(z,c)+b_1^d)\ldots) + \begin{pmatrix}b^{\mu'}\\ b^{\sigma'}\end{pmatrix}, \tag{4.2a}$$

$$\hat{x} = \mu' + \epsilon \odot \sigma'\,, \tag{4.2b}$$

where W_l^d and b_l^d denote weight matrices and bias vectors for decoding, respectively. μ' and σ' parameterize a z-dependent normal distribution in the x space.

[1] In contrast with Chapter 2 and Chapter 3, x are data space vectors, and z are latent space codes.

4.2.2. TRAINING AND GENERATION PROCESS

In the training stage, weight matrices W and bias vectors b are updated in an iterative way with the goal of minimizing the loss function [92]

$$\mathscr{L} = \mathscr{L}_{D_{KL}} + \mathscr{L}_{Re}. \tag{4.3}$$

The *Kullback-Leibler loss* $\mathscr{L}_{D_{KL}} = \sum_i D_{KL}(q_\phi(z|x_i)||p(z))$ is the sum over all training data points x_i (assumed i.i.d.) of the Kullback–Leibler divergence between that point's posterior distribution $q_\phi(z|x_i)$ and the prior distribution $p(z)$ (chosen as the standard normal distribution). The posterior distribution $q_\phi(z|x_i)$ is determined by the parameters ϕ of the encoder network and represents the mapping of the point x_i into a normal distribution in the latent space using (4.1a) and (4.1b). As the Kullback-Leibler divergence between two normal distributions can be evaluated directly [88], the Kullback-Leibler loss is computed as

$$\mathscr{L}_{D_{KL}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{d}(-1+\sigma_{i,j}^2+\mu_{i,j}^2-\log\sigma_{i,j}^2), \tag{4.4}$$

where n denotes total number of observations used for training and (μ_i, σ_i) are evaluated for data point x_i and condition c using (4.1a).

The *reconstruction loss* $\mathscr{L}_{Re}$ stands for the negative log-likelihood of reconstructing the inputs x_i via their latent space codes and the decoder that is parameterized by θ. The reconstruction loss is thus computed as

$$\begin{aligned}\mathscr{L}_{Re} &= -\sum_{i=1}^{n}\mathbb{E}_{Z\sim q_\phi(z|x_i)}[\log p_\theta(x_i|Z)] \\ &\approx \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{d}((x_{i,j}-\mu'_{i,j})^2/\sigma'^2_{i,j}+\log\sigma'^2_{i,j})+\frac{nd}{2}\log 2\pi,\end{aligned} \tag{4.5}$$

where the final step involves a single-point approximation of the expectation and (μ'_i, σ'_i) are obtained from the randomly generated latent code $z(x_i)$ and the condition c using (4.2a). During training, the full-sample sum in loss functions (4.4) and (4.5) are replaced by batch-sample averages. The constant $\frac{nd}{2}\log 2\pi$ of $\mathscr{L}_{Re}$ is omitted.

After the training process, only the decoder part of the trained CVAE network is utilized to generate data. Latent space codes $\tilde{z}$ are sampled from the standard normal distribution $\mathscr{N}(0, I)$ (see Fig. 4.1b). Then, data space samples $\tilde{x}$ are sampled from distribution $\mathscr{N}(\mu'(\tilde{z}, c), \sigma'(\tilde{z}, c))$, whose parameters are determined by $\tilde{z}$ and c using (4.2a). We note that although the amount of available training data determines the information

contained within the model, there is no limit to the amount of data that can be generated.

In this way, a complex data distribution in the x space is constructed as a continuous superposition of normal distributions that are parameterised by the normally distributed coordinate z. Using the procedure above, the encoder and decoder networks are trained to adapt any distribution to this normally distributed latent space. We note that other distributions besides the normal distribution can be used as the prior for the latent space coordinate z [93], [94] – selecting the best latent space representation for a particular class of problems remains an open research problem.

4.2.3. GENERATOR OPTIMIZATION STRATEGY

NETWORK AND OUTPUT NOISE CO-OPTIMIZATION STRATEGY

It is common for CVAE implementations to generate data $\tilde{x}$ not by sampling from the distribution $\mathcal{N}(\mu'(\tilde{z},c),\sigma'(\tilde{z},c))$ via (4.2b), but by directly using the mean value $\mu'(\tilde{z},c)$ (the maximum likelihood sample). Moreover, the standard deviation σ' is not co-optimized in the training process of (4.3), but considered a hyperparameter that fixes $\sigma'_{i,j}$=s identically in all dimensions, so that (4.5) can be replaced by

$$\tilde{\mathscr{L}}_{Re} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{d}\frac{(x_{i,j}-\mu'_{i,j})^2}{s^2}. \tag{4.6}$$

In contrast, we investigate the model in which the parameters σ' of the output noise distribution are co-optimized as a function of z during training, as was recently also (independently) proposed in [87]. In addition, we explicitly add output noise $\epsilon\odot\sigma'(\tilde{z},c)$ to the generated data. To compare the different approaches, the quality of the generated data is evaluated under all four combinations (Table 4.1): whether σ' is co-optimized in the training stage (Auto σ') or set to a fixed value (Fixed σ'); whether the noise $\epsilon\odot\sigma'(\tilde{z},c)$ is added to the outputs (Noisy) or not (Noise free).

LOSS FUNCTION WEIGHT TUNING STRATEGY

The two loss terms have opposing effects. The Kullback-Leibler loss $\mathscr{L}_{D_{KL}}$ ensures a good fit with the prior distribution that samples are generated from, thus suppressing spurious generated points at the expense of 'smoothing' the output. The reconstruction loss $\mathscr{L}_{Re}$, on the other hand, promotes exact reconstruction of the training data. In this section, in addition to the output noise, we also study the effect of a heuristic weight-

Table 4.1: Overview of model permutations used in experiments

Strategy (with β)	**Objective function**	**Generation**
Auto σ', Noisy	$\beta\mathscr{L}_{D_{KL}} + \mathscr{L}_{Re}$	$\mathscr{N}(\mu'(\tilde{z},c),\sigma'(\tilde{z},c))$
Auto σ', Noise free	$\beta\mathscr{L}_{D_{KL}} + \mathscr{L}_{Re}$	$\mu'(\tilde{z},c)$
Fixed σ', Noisy	$\beta\mathscr{L}_{D_{KL}} + \tilde{\mathscr{L}}_{Re}$	$\mathscr{N}(\mu'(\tilde{z},c), sI)$
Fixed σ', Noise free	$\beta\mathscr{L}_{D_{KL}} + \tilde{\mathscr{L}}_{Re}$	$\mu'(\tilde{z},c)$

ing factor β [95] for the Kullback-Leibler loss term $\mathscr{L}_{D_{KL}}$ (so-called β-VAE) on statistical properties of the generated data. It is written as

$$\mathscr{L} = \beta\mathscr{L}_{D_{KL}} + \mathscr{L}_{Re}. \tag{4.7}$$

All aforementioned combined strategies are explicated in Table 4.1. Their impacts on the quality of generations will be investigated in the following sections. Particularly, the settings of standard deviation σ' and weight β influence the objective function in the training process and will ultimately affect the generated data. On the other hand, the use of output noise, $\epsilon \odot \sigma'(\tilde{z},c)$, will directly impact the data generation stage.

4.3. CASE STUDY ON COUNTRY LEVEL LOAD DATA

To validate a generative model, an important aspect is to evaluate the quality of the generated data. To do so, not only visual comparison of the similarity of generations and historical data but also statistical assessments are expected. In this section, the performance of our proposed CVAE-based generative model is analyzed using a European load data set. This is done with three data quality metrics that measure univariate distributions and multivariate dependencies. Impacts on the quality of generated data are investigated under the experimental settings in Table 4.1, using both conditional and regular VAEs. The model performance under different weighting factors β is tested, and different generative models are compared.

4.3.1. DATA SOURCE AND GENERATION

Historical hourly load data for 32 European countries between 2013 and 2017 was obtained from the Open Power System Data platform [48] (package version *2019-06-05*). Columns of AL (Albania), CS (Serbia and Montenegro), CY (Cyprus), GB (United King-

dom), TR (Turkey) and UA (Ukraine) were dropped for incomplete records. The historical data were randomly split into training and test sets in blocks of one week with a proportion of 4:1 (35,148 training and 8,569 test samples). The training set was min-max normalized before being fed into the CVAE model and the inverse transformation was applied to generate samples. The contextual information c is the hour of day. Both total and hourly volumes of the generated data are the same as the training data set, in order to balance them for visual and statistical analysis. However, we emphasize that the purpose of constructing such a generative model is to have the ability to generate limitless non-repeating data, e.g., for reducing the risk of overfitting in downstream machine learning tasks.

The parameters of the generative models were tuned for optimal performance, for both the VAE and the CVAE. The network contained 2 hidden layers in the encoder with dimensions of 24 and 16, respectively; the bottleneck layer had 8 nodes (8-dimensional latent vector). The decoder also had 2 hidden layers with the same dimensions in reverse order. Comparisons against 4-neuron and 16-neuron bottlenecks revealed that a smaller bottleneck results in excessive loss, whereas a larger bottleneck insufficiently forces the network to learn features. In the CVAE model, the hourly time-of-day was encoded cyclically using sine/cosine representation.

The ReLU activation function was used, except for the generation of μ and σ leading up to the bottleneck and output layers. The *adaptive moment estimation* (Adam) weight optimizer [51] was utilized with default settings to iteratively optimize the value of weight matrices W and bias vectors b. The batch size and learning rate related parameter α for training was 64 and 10^{-4} respectively and 20,000 training iterations were used. Training and data generation of the model was conducted in Python using `tensorflow` on the Google Colab environment using the GPU option. The code used for this chapter is available for download[96], [97].

4.3.2. DATA QUALITY METRICS

To test a generative model's ability to reproduce the features of historical data, especially in high dimensions, statistical tests are required. Three tests are put forward to examine different aspects of the generated data set, in comparison with the historical data.

KOLMOGOROV-SMIRNOV TEST FOR UNIVARIATE MARGINAL DISTRIBUTIONS

We used the two-sample Kolmogorov-Smirnov (K-S) test [98] to see whether the generated data was able to reproduce the marginal load distributions for each of the countries in the data set. For a given output dimension (load in a single country), historical and generated data are compared. Under the null hypothesis that historical and generated data are drawn from the same model, the p-values should follow a uniform distribution. In other words, when the historical data is compared against itself, the cumulative distribution of p-values should lie on the diagonal. Thus, for generative models, the closer the cumulative distribution of p-values lies to the diagonal, the higher the similarity between the two distributions.

Clearly, the models are unlikely to exactly reproduce the historical distribution, thus large deviations from the ideal curve will show up for large-sample tests. Nevertheless, to analyze the degree of performance of various models, we use repeated tests on smaller sample sets that result in clear differentiation, as in [78]. In this chapter, 0.5% of the data set, i.e. 176 data points out of 35,148, were randomly drawn from training and generated data set, and then a p-value was calculated accordingly. This process was repeated 5,000 times for each country and a curve was constructed from all p-values.

AUTOENCODER-BASED POINT-WISE TEST FOR MULTIVARIATE DEPENDENCIES

Autoencoder (AE) neural networks have been proven to be highly sensitive anomaly detectors in Chapter 2 and Chapter 3. Unlike (C)VAE networks, AEs have no stochastic layers and only minimize the reconstruction loss $r = \sum_i \|x_i - \hat{x}_i\|^2/d$. An AE learns to compress and decompress the data based on properties of the training set. As a result, data points with dependencies that deviate substantially from that in the training set tend to have higher reconstruction errors.

An independent AE network was trained for this test, with hyperparameters equal to that of the CVAE model, except for the stochastic layers and objective function. Reconstruction errors of all data points (historical or generated) are plotted as cumulative distributions for easy comparison. As a test for overfitting of the autoencoder on the training data, the autoencoder test was performed on the training and test data. The two distributions visually overlapped, suggesting this is not a concern.

ENERGY TEST FOR MULTIVARIATE DEPENDENCIES OF POPULATION

Another two-sample test, the energy test [99], was conducted to examine whether the *multivariate dependencies* of the population were well acquired from the training set.

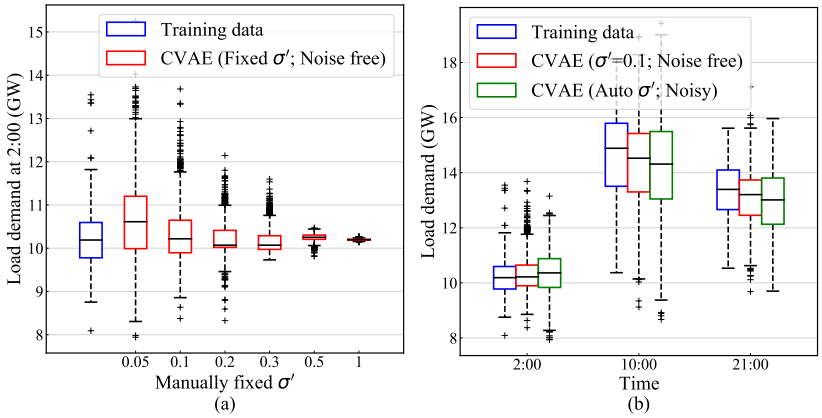

Figure 4.2: (a) Box plots of the original and generated load values in the Netherlands, based on 1459 data points at 2:00. Data was generated by CVAE using 'Fixed σ', Noise free' approach with different σ'. (b) Distribution comparison of the original and generated load data in the Netherlands, based on 1459 (2:00), 1465 (10:00) and 1465 (21:00) data points, respectively. Data was generated by CVAE using 'Fixed σ', Noise free' method (σ'=0.1) and 'Auto σ', Noisy' scheme.

The energy test, computed using the PyTorch library *torch-two-sample* [100] uses a user-specified number of permutations (200 was used) to calculate a p-value. As for the energy test, we used random subsets of 0.5% of the generated population and historical population. We repeated this process 1,000 times to draw a distribution of p-values and compare it with the uniform distribution (which would be expected if the data was drawn from the historical distribution).

4.3.3. EXPERIMENTAL RESULTS ANALYSIS

VISUAL COMPARISON OF UNIVARIATE DISTRIBUTIONS

In this first experiment, the CVAE with fixed σ' and no output noise was used to generate 1,459 load demands, conditioned on the time 2:00. Results for the Netherlands are shown in Fig. 4.2a, for various values of σ'. As the output noise assumed in training increases, the variability of the generated points decreases (because noise is not actually added). When $\sigma' = 0.1$, the distribution of generations is the closest to that of historical data. This setting will be used for all further experiments with fixed σ'.[2]

Fig. 4.2b further compares data generated using the 'Fixed σ'' and 'Auto σ', Noisy'

[2]Note that we only fix a single parameter in this case. An approximate visual match of the box plots is a necessary condition for a good overall fit, justifying the choice $\sigma' = 0.1$ for this comparison.

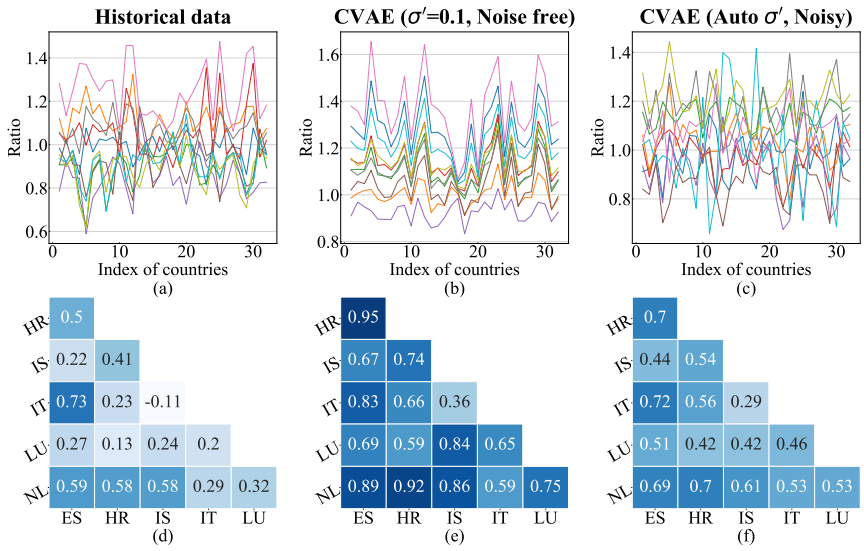

Figure 4.3: (a), (b), and (c) display 10 typical ratios of 32 countries' historical and generated data to the historical mean values at 2:00. (d), (e) and (f) demonstrate the Pearson correlation coefficient matrices of 6 (out of 32) countries' historical and generated data at 2:00. The horizontal and vertical dimensions in the matrices are Spain (ES), Croatia (HR), Iceland (IS), Italy (IT), Luxembourg (LU) and the Netherlands (NL).

schemes and the training data. Conditioning on 2:00, 10:00, and 21:00 was performed, and results are shown for the Netherlands. Both methods are able to qualitatively reproduce the features of the data.

MULTIVARIATE CORRELATIONS

The top row of Fig. 4.3 shows the loads of all countries for 10 different snapshots at 2:00, relative to the mean load in those countries at 2:00. Compared to historical data (a) and the noisy generator (c), samples generated by the noise-free generator clearly show higher correlations between countries. This is confirmed by the correlation analysis between six countries in the bottom row of Fig. 4.3. By omitting output noise, the noise-free generator generated (too) highly correlated samples.

Sensitive experiments for the multivariate dependencies will be conducted in the following sections using the autoencoder-based point-wise test. The accurate representation of multivariate dependencies will be important for the analysis of supply shortfalls in Section 4.4.

Influence of noise generation

In this experiment, the influence of the four strategies listed in Table 4.1 were tested with $\beta = 1$. Results for the statistical tests described in Section 4.3.2 are shown in the first row of Fig. 4.4. The K-S test results and autoencoder results show that the inclusion of output noise is essential to improve marginal distributions (Fig. 4.4a) and increase output variability to the level of the historical data (Fig. 4.4b). In addition, the autoencoder and energy tests show that automatic tuning of the noise strength (Auto σ') is essential to improve the multivariate dependencies of the generated samples. Together, this experiment shows that the 'Auto σ', Noisy' generator outperforms the other approaches listed in Table 4.1. This was to be expected given the mathematical theory behind the CVAE (which includes noise), but is at odds with common implementations.

Comparison between conditional and regular VAE

In the second row of Fig. 4.4, the performance of the CVAE and VAE models (with $\beta = 1$) is compared. The CVAE model slightly outperforms the VAE model in all categories. One possible explanation is that the CVAE model has access to the context c (time of day), which effectively increases the dimension of the latent space. Because of its better performance, we continue using the CVAE model in subsequent experiments, but the results suggest that a VAE model delivers comparable performance, and may be preferable when no natural conditioning variable is available.

β sensitivity test

The third row of Fig. 4.4 shows the impact of β (values 1, 3, 10) on the performance of the CVAE (Auto σ', Noisy) model. As β is increased, the performance on the K-S test (Fig. 4.4g) improves, indicating an improved ability to learn marginal distributions. On the other hand, performance on the autoencoder test (Fig. 4.4h) worsens, suggesting that points 'outside' of the training point cloud are generated for large β. Finally, the energy test (Fig. 4.4i) indicates that a moderate value of β can strike a balance between the opposing requirements: the curve for $\beta = 3$ is closest to the desired result. Nevertheless, depending on the application, it may be desirable to choose β larger or smaller.

Comparison of Generative Models

In the fourth row of Fig. 4.4, the quality of data sampled from different generative models was investigated. The values of β for CVAE and VAE models (both Auto σ', Noisy)

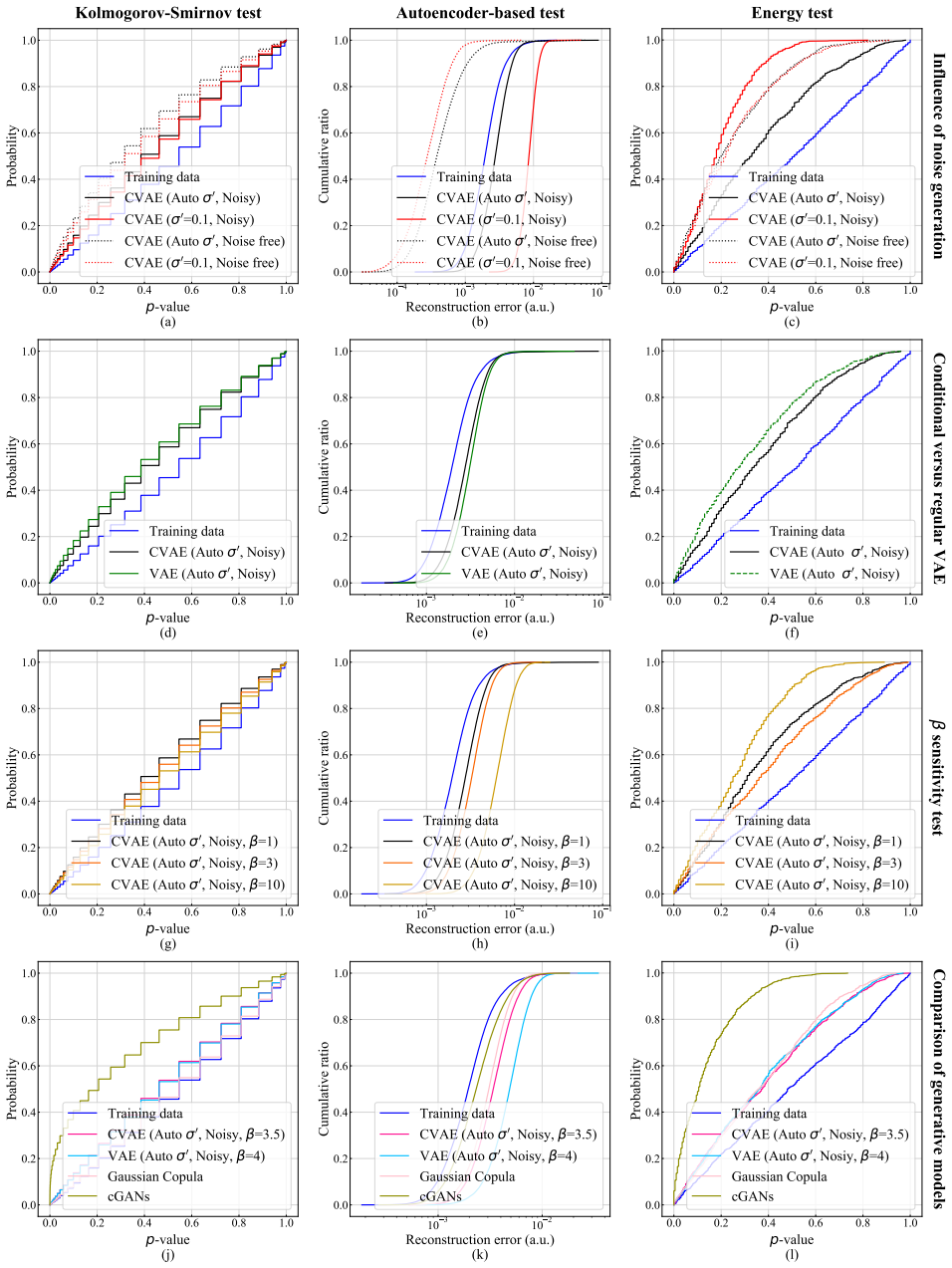

Figure 4.4: Results of statistical tests. Each row denotes a set of experiments (noise generation, training condition, value of β and model family). The three columns depict results for the three tests described in Section 4.3.2.

were tuned for optimal performance on the energy test (see previous section). In addition, *Gaussian copula* [31] and cGAN [101] models were included for comparison. The basic cGAN model was modified to use Wasserstein losses[102]. Both its generator and discriminator are deep neural networks; each has two hidden layers of 256 neurons, all activated with LeakyReLU (α = 0.2) except in the output layers, where linear and sigmoid activation functions are used for the generator and discriminator. Weights of the neurons are optimized with *root mean square propagation* (RMSprop) available from the python package Keras.

The K-S test shows the outstanding ability of the Gaussian copula model to reproduce marginal load distributions (a design feature of copula models [78]). This model also shows competitive performance on the autoencoder and energy tests. However, it will become clear in Section 4.4 that its multivariate tail performance is worse than that of the (C)VAE models. The cGAN model shows the best performance on the autoencoder test, indicative of its ability to generate samples with realistic features. However, the model significantly underperforms on the K-S and energy tests, which suggests that the generated samples, though 'realistic', are unevenly distributed through the space of possible states. The optimized CVAE and VAE models are competitive on all three tests, with the CVAE model slightly outperforming the VAE model.

4.4. CASE STUDY ON MULTI-AREA ADEQUACY ASSESSMENT

Evaluating the potential risk of grids through enormous generated representative scenarios is valuable for power system stable operation, especially when the data are not sufficient. In this section, we investigate the performance of the load generation mechanisms by using it for a multi-area adequacy assessment study, based on the ENTSO-E Mid-term Adequacy Forecast 2020 (MAF2020) [103]. Multi-area adequacy assessment measures the sufficiency of generating capacity compared with the load on each of the nodes in the power system under transmission constraints. This can be considered a stress test of the generative model, as the outcomes are sensitive to high-load events (tail distributions) *and* their dependencies between countries.

Loss Of Load Expectation (LOLE [h/year]) and *Expected Energy Not Served* (EENS [MWh/year]) were estimated by Monte Carlo simulations. LOLE is the expected number of hours per year during which the supply does not meet demand. EENS is the expected amount of energy demand per year that cannot be supplied. Parameters from the

MAF2020 study were used to construct a model for generating capacity and net transfer capacities between countries. They were combined with generated and historical load data to define a probabilistic model for the Monte Carlo simulations. We emphasize that the model thus constructed is not meant to be an accurate representation of the European grid, but a stylized problem that serves as a comparative testing ground for the generative models.

4.4.1. MULTI-AREA ADEQUACY ASSESSMENT STRUCTURE

We consider the network as a directed graph (to allow for asymmetric flow limits) where nodes are zones, edges are connections between zones, and edge capacities are transfer capacities. Each sampled state w is represented by the available generating capacity $\overline{g}_i^w$ and demand d_i^w of each node i. Based on the flow constraints and dispatching policy, the consumed power p_i^w and load curtailment c_i^w for each node can be calculated, related by

$$c_i^w = \max(0, d_i^w - p_i^w). \tag{4.8}$$

We determine c_i^w (and implicitly p_i^w) by solving a quadratic problem with variables $\tilde{c}_i$ (curtailment) and $\tilde{f}_{ij}$ (flows), which aims to minimize the total load curtailments and assumes that curtailments are balanced between areas [104], relative to the demand in that area:

$$\vec{c}^w = \arg\min_{\tilde{f},\tilde{c}} \sum_{i\in\mathcal{N}} \frac{1}{2d_i^w}\tilde{c}_i^2 + \tilde{c}_i \tag{4.9}$$

$$\underline{f_{ij}} \le \tilde{f}_{ij} \le \overline{f_{ij}}, \qquad \forall (i\,j) \in \mathcal{L} \tag{4.10}$$

$$0 \le \tilde{c}_i \le d_i^w, \qquad \forall\, i \in \mathcal{N} \tag{4.11}$$

$$d_i^w - \overline{g}_i^w \le \sum_{j<i} \tilde{f}_{ji} - \sum_{j>i} \tilde{f}_{ij} + \tilde{c}_i \le d_i^w, \qquad \forall\, i \in \mathcal{N} \tag{4.12}$$

Here, $\mathcal{L}$ and $\mathcal{N}$ are the sets of connections (from i to j with $i < j$) and areas respectively. Constraints on power flow $\tilde{f}_{ij}$ from node i to node j are given in (4.10); (4.11) limits curtailment and (4.12) enforces flow and generating power constraints. The objective function (4.9) has a positive definite structure and the constraints are linear, so this optimization problem is strictly convex and has a unique solution. This optimization problem was solved using the python package *quadprog* [105].

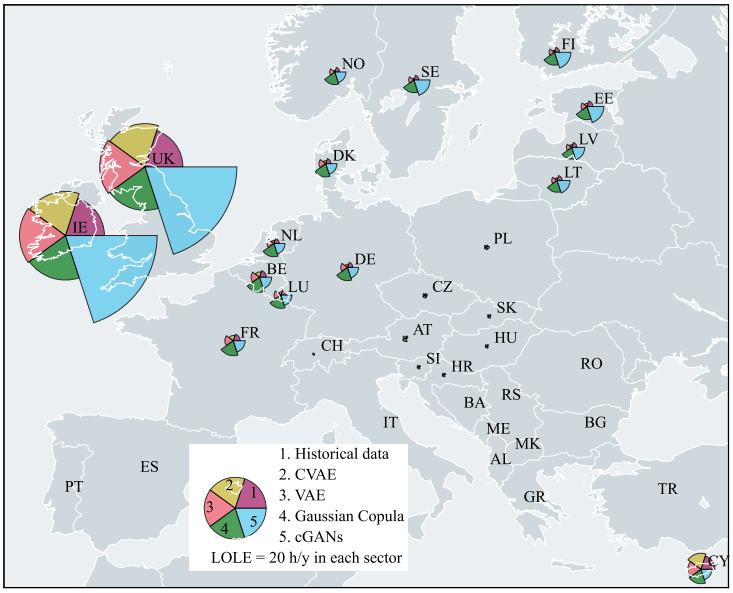

Figure 4.5: Comparison of LOLE estimates using historical load data and generative load. The area of each sector of the disc represents the LOLE of the corresponding model (20h/y shown for scale in legend).

4.4.2. POWER SYSTEM MODEL

A European adequacy assessment model was developed, based on the target year 2025 data from the ENTSO-E MAF2020 [103]. The net transfer capacities between countries are defined as the summation of transfer capacities between their constituent zones, as reported in the MAF2020. Since details of generators and unit capacities are not reported in the released dataset, we model the total generating capacity and the unit capacities in each country as follows. The assumed generating capacity of each country is a summation of conventional generating capacity in its zone(s) plus 5% of nameplate wind power capacity. Unit sizes are set per country as the closest value under 500MW that is a divisor of the generating capacity; a unit availability of 83% is used. Cyprus has no connection to other countries, so a unit capacity of 95MW is used to avoid excessive outages.

4.4.3. MULTI-AREA ADEQUACY ASSESSMENT RESULTS

To compare the CVAE, VAE, Gaussian Copula, and cGAN generators, they were trained on historical load data from 2017 and 2018 for 35 countries, retrieved from the Open Power System Data Platform ([48]; columns for CS, IS, and UA were omitted). Each model was

used to generate 100,000 random load samples. The 'Auto σ', Noisy' setting was utilized for the CVAE and VAE models, and β was set to 10 for improved reproduction of the marginal distributions. For each model, 1,000,000 Monte Carlo generation samples were drawn and combined with random load samples to estimate the LOLE in each country. Fig. 4.5 depicts the estimated LOLE values for all generative methods and historical data. The area of each sector of the disc represents the LOLE obtained using a particular load states generating model. Numerical results for three countries with low (Austria, AT), medium (The Netherlands, NL), and high (UK) risk levels are shown in Table 4.2. Standard errors for the least significant digits are shown in parentheses. Moreover, to investigate the beneficial effect of interconnection – and therefore the importance of accurate multivariate modeling – risks for these countries are also reported in the absence of interconnection ('island').

By design, the Gaussian copula reproduces the marginal distributions of the historical data. Therefore, the calculated risk for *islanded* systems is consistent with those for historical data. However, the results demonstrate that this model tends to overestimate risks for interconnected nodes (countries). The cGAN generative model tends to cause an overestimation of risks with *both* islanded and interconnected nodes, sometimes very significantly (e.g. the LOLE values for the UK and Ireland). In comparison, both the VAE and CVAE models generate data that results in risk estimates that are closer to those observed using historical data, although deviations exist from country to country. This suggests both models are able to substantially represent the multivariate tail distribution of the historical data.

The capability of generating load conditioned on hours is an additional advantage of CVAE in comparison with VAE in the adequacy assessment context. Load curtailments usually accrue during high load hours. So, time of day could be used as a control variable for an importance sampling Monte Carlo scheme that preferentially samples load states at high load hours and compensates for the resulting bias by sample re-weighting.

4.5. CASE STUDY ON LOAD DATA OF INDIVIDUAL CUSTOMERS

In section 4.3, the impact of the CVAE model's output noise on its generative performance has been investigated with a use case of learning and generating snapshots of country-level load states. However, such snapshots of large load aggregations have limited diversity and variability.

Table 4.2: Calculated risks of selected countries (with and without interconnection) using historical data and all generative models.

Country	LOLE (h/y)					EENS (MWh/y)				
	historical data	CVAE	VAE	Gaussian copula	cGAN	historical data	CVAE	VAE	Gaussian copula	cGAN
AT	0.03(2)	0(0)	0.18(4)	0.32(5)	0.05(2)	4(2)	0(0)	33(11)	42(10)	10(5)
NL	0.74(8)	0.12(3)	0.80(8)	4.6(2)	3.6(2)	119(17)	33(11)	300(45)	1135(67)	1743(129)
UK	37.8(6)	50.6(7)	54.1(7)	50.2(7)	223.6(14)	$5.20(10)\cdot10^4$	$8.77(15)\cdot10^4$	$1.16(2)\cdot10^5$	$7.75(14)\cdot10^4$	$4.99(4)\cdot10^5$
AT (island)	0.74(8)	1.01(9)	0.88(9)	0.80(8)	0.61(7)	221(33)	435(54)	334(45)	273(40)	255(41)
NL (island)	63.8(7)	65.2(8)	69.2(8)	69.4(8)	99.7(9)	$4.13(7)\cdot10^4$	$4.35(7)\cdot10^4$	$4.73(7)\cdot10^4$	$4.61(7)\cdot10^4$	$6.74(8)\cdot10^4$
UK (island)	1026(3)	982(3)	884(3)	1033(3)	1965(4)	$4.28(2)\cdot10^6$	$4.24(2)\cdot10^6$	$3.88(2)\cdot10^6$	$4.38(2)\cdot10^6$	$1.078(3)\cdot10^7$

This section investigates the CVAE model's capacity to generate synthetic load profiles that are representative of those from a large variety of individual users. Compared to section 4.3, this section aims to generate consumption patterns that are temporal (instead of spatial) and at a lower aggregation level, where the loads are more stochastic. To do so, we first analyze the properties of daily load profiles of an anonymized data set of 5,000 industrial and commercial customers. Moreover, for better training and generation performance, we introduce data split, month condition, and power exchange intensity calculation strategies during data processing. Eventually, we evaluate the performance of the CVAE model under different time (month) and power exchange intensity conditions with both visual and statistical metrics.

4.5.1. DATA SOURCE AND GENERATION

We used the CVAE-based generative model described above to generate daily load profiles (24 hours) of individual network connections (i.e., users), conditioned on the month of the year and power the user typically exchanges with the grid. The performance of the model was analyzed using a load data set of 5,000 users. The quality of generations was evaluated visually as a function of conditioning parameters. In addition, performance was validated statistically by measuring univariate distributions and multivariate dependencies. Moreover, an experiment was conducted to test the model capacity of interpolation.

Anonymized historical electricity consumption/generation data of 5,000 industrial and commercial electricity users during 2020 was obtained from Alliander NV[106], a Dutch distribution network owner and operator. The time resolution of the data is 15 minutes. It is worth noting that the data set's time label is *UTC* (Coordinated Universal Time). However, the actual local time for electricity users is *CET* (Central European Time). During standard time and daylight saving time, their time differences are 1 and 2 hours, respectively. The energy data was converted from integer kWh values to average power with multiples of 4 kW. Compared with country-level load profiles[48], the energy consumption of individual users involves more variability and less predictability. Fig. 4.6 illustrates the large variety of daily profiles, by plotting the marginal histogram and joint density of all historical load profiles at 10:00 and 21:00. Note the logarithmic density used, indicating a large concentration around (relatively) small values. Moreover, data points located in the upper-left and bottom-right quadrants stand for users that can not

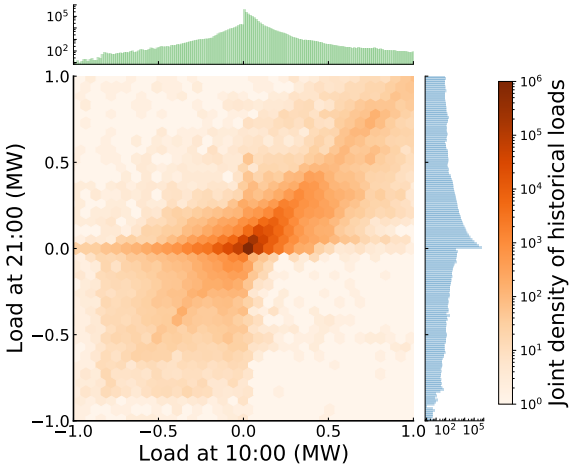

Figure 4.6: Marginal histogram and joint density of historical loads at 10:00 and 21:00 during one year.

only consume but also generate energy. All these factors above make it challenging for the CVAE model to capture the load patterns.

The data processing scheme is shown in Fig. 4.7. The historical data were split, scaled, and conditioned. Three data process strategies used in this study are as follows.

STRATEGY I - DATA SPLIT

The historical data were randomly split into training and test sets as blocks of one week with a proportion of 4:1. This strikes a balance between separating individual days (subsequent days are not sufficiently independent) and separating larger blocks (insufficient coverage in the test set).

STRATEGY II - MONTH CONDITIONS

In this study, one of the conditions (contextual information) c is the month of the year. We used the $\sin(\cdot)$ and $\cos(\cdot)$ values of a month as the condition of load data. For a specific month m, its condition was encoded as $\sin(\frac{m}{12} \cdot 2\pi)$ and $\cos(\frac{m}{12} \cdot 2\pi)$. This encoding reflects the continuity and circularity of this feature.

STRATEGY III - USER INTENSITY

After inspecting historical load profiles, we noticed that some users had relatively regular load profiles, whereas others had irregular behavior with rare consumption or generation spikes. Some connections were only active during a small part of the year. To

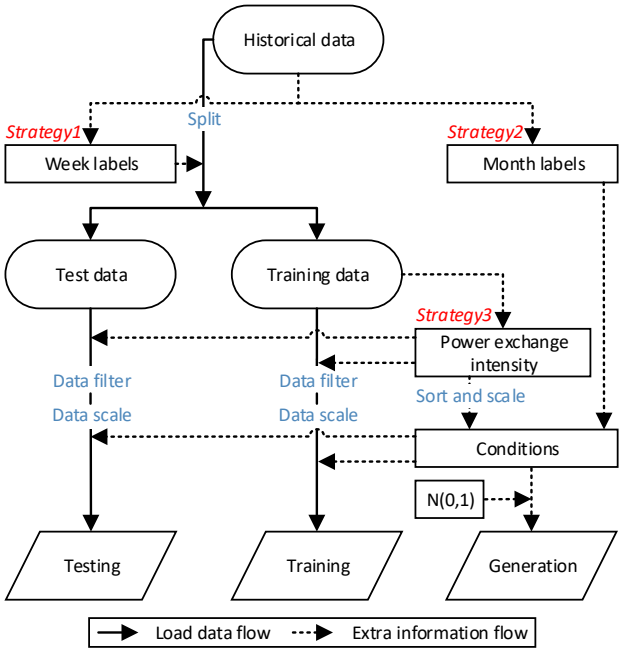

Figure 4.7: Data processing scheme.

construct a conditioning feature that represents the 'size' of electricity users, we calculated the daily average power exchange by averaging over all non-zero values of the *absolute* power (consumption or generation) for each user and each day. For each customer, this value was averaged over the five days with the largest daily average power exchange to obtain the *user intensity*. The intensity values were used to assign to each customer a rank order $c \in [0,1]$. Due to the large range of power values present in the data (see Fig. 4.6) and the relative scarcity of data with high peak exchange, we trained the model only on profiles of customers with an intensity up to 100 kW. Ultimately, 4,049 users remained, with 1,170,110 and 307,720 load profiles in the training and test sets, respectively. The values were scaled by $1/(100\ kW)$ for training.

The parameters of the generative models were tuned for optimal performance. The input and output layers had 96 dimensions (24 hours with 15-minute resolution). Accordingly, 96-dimensional daily load profiles were used for training and generation. The network contained 3 hidden layers in the encoder with dimensions of 800; the bottleneck layer had 12 nodes (12-dimensional latent vector). The decoder also had 3 hidden layers with dimensions of 800. The contextual condition c consisted of a 2-dimensional

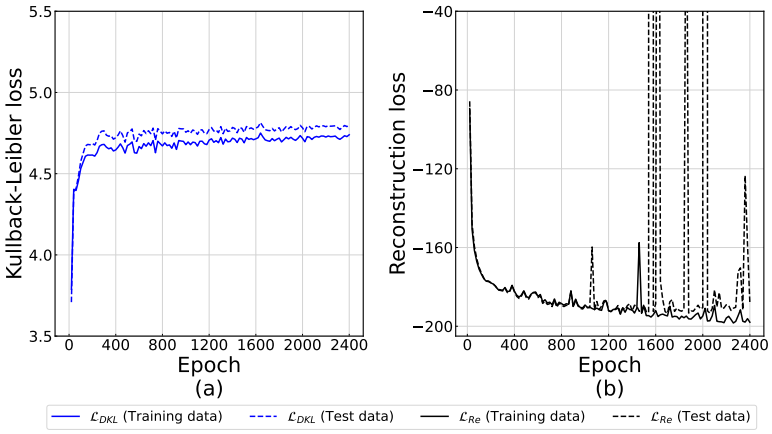

Figure 4.8: Training process and its failure. (a) Training process of Kullback-Leibler loss. (b) Training process of reconstruction loss.

month condition and a 1-dimensional per-user power exchange intensity.

The ReLU activation function was used, except for the generation of μ (μ') and σ (σ') leading up to the bottleneck and output layers. The *adaptive moment estimation* (Adam) weight optimizer [51] was utilized with default settings to iteratively optimize the value of weight matrices W and bias vectors b. The batch size and learning rate parameter α for training were 1,280 and 10^{-5} respectively. The weighting factor β was set as 8.5. Training and data generation of the model was conducted in Python using `tensorflow` on the Google Colab environment using the GPU option. The training process is shown in Fig. 4.8. The Kullback-Leibler loss rapidly stabilizes during training. However, the reconstruction loss of the test data set starts to deviate from the training loss and fluctuates strongly after 1,000 training epochs, which indicates an overfitting of training data and general training instability. To find a compromise between loss minimization and generalization capacity of the trained model, 1,000 training epochs were used in this research. During the generation process, the total, monthly, and per-user amounts of synthetic data are identical to the training set.

4.5.2. EXPERIMENTAL RESULTS ANALYSIS

COMPARISON OF DAILY LOAD PROFILES

To validate the generation capacity of our proposed CVAE model, we first visually inspected the generated contextual load profiles. We defined the customers with the first

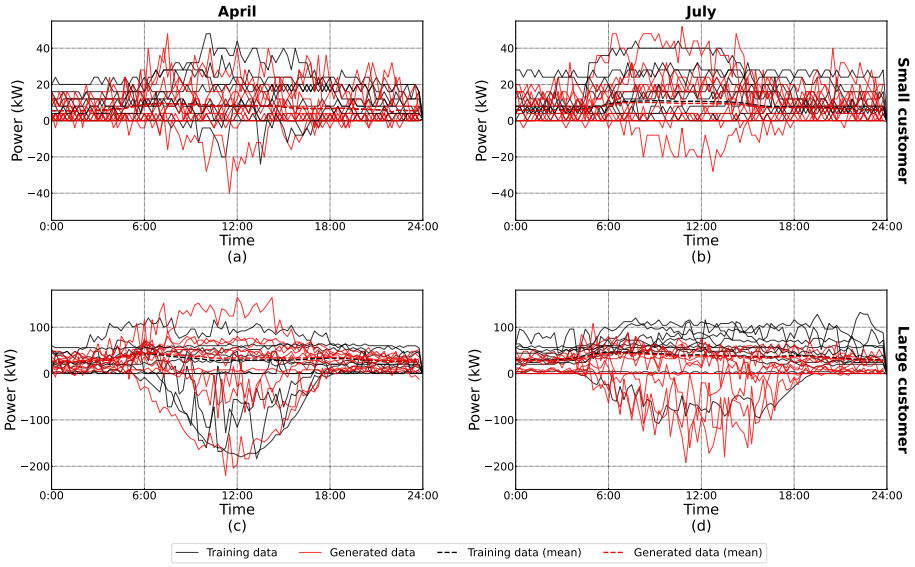

Figure 4.9: 10 randomly sampled historical and generated load profiles and average daily load from customers of various sizes in different months.

and last 30% of per-user intensities as small and large customers, respectively, and the remaining 40% of users as medium users. In this experiment, we conditioned the generation of profiles on the months of April and July, and the 'small' and 'large' customer classes (by random sampling of c in their respective ranges). Fig. 4.9 shows the mean value of load profiles under each condition combination (generated versus measured), and 10 randomly sampled load profiles alongside 10 random historical profiles. The mean generated load under each condition combination has a similar curve shape to the training data. Moreover, compared with historical data, the displayed load generations retain randomness and show a sense of realism, indicating that the CVAE model captures temporal features of historical load profiles.

CLUSTERING PERFORMANCE

The following experiment compared all historical and generated daily load profiles for a more elaborate test of the distribution of generated load profiles. We first split the training data set into 8 clusters by the K-means algorithm [107], using the squared Euclidean distance metric. Then, we assigned the generated and test load profiles to the nearest cluster. The mean values of training, test, and generated loads for each cluster are depicted in Fig. 4.10a-h, in decreasing order of training data volume. The most voluminous

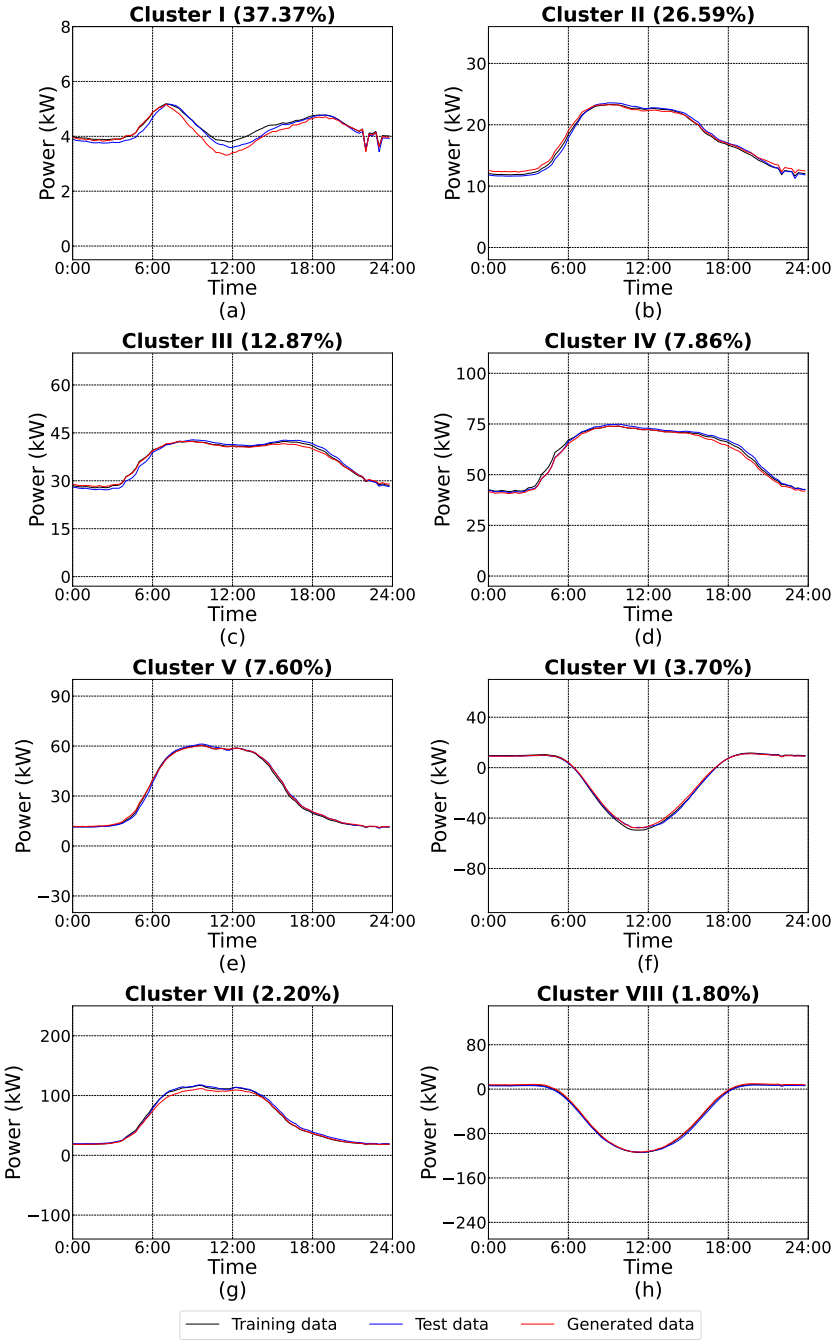

Figure 4.10: Mean value of historical and generated data in different clusters.

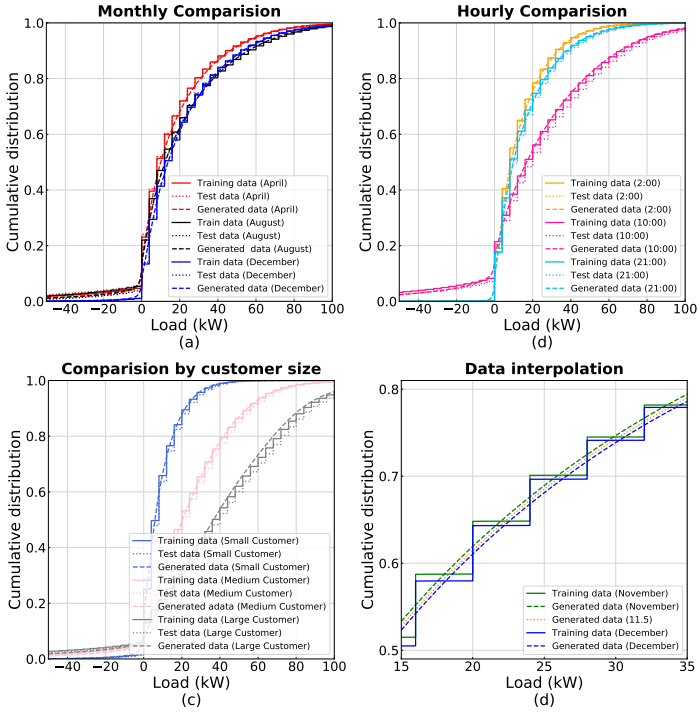

Figure 4.11: Cumulative distribution comparison of historical and generated data via different time scales and for users of various sizes.

cluster has small average load values. Note that the apparent gap in cluster I is smaller than the resolution of the data. Some clusters correspond to larger loads and generators (mainly solar PV). In all cases, the mean values of profiles assigned to the cluster match well.

MARGINAL DISTRIBUTION COMPARISON

The third experiment compared the cumulative distribution of historical and generated data via different time scales and users of various sizes. The experimental results are shown in Fig. 4.11; note the discretization of the real measurements, visible in these graphs. Fig. 4.11a exhibits the cumulative distribution of loads in different months. The CVAE model is able to generate contextual load profiles that follow the monthly distribution variation of historical loads. The hourly comparison of the load depicted in Fig. 4.11b shows that the curves of generations overlapped with the historical training data, demonstrating quite similar hourly distributions. Moreover, the comparison re-

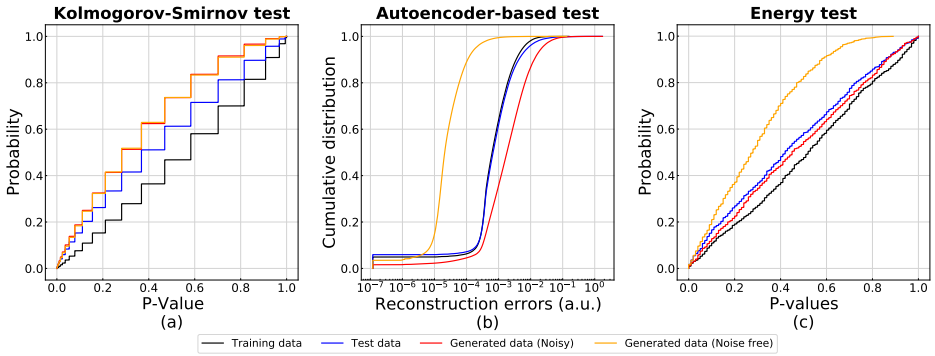

Figure 4.12: (a) Kolmogorov-Smirnov test. (b) Autoencoder-test. (c) Energy test

sult shown in Fig. 4.11c stands for a good capture of load patterns of different customer sizes. Finally, we tested the interpolation capacity of the CVAE model. Specifically, we used a virtual month condition (11.5) to generate load profiles, and the result is shown in Fig. 4.11d. The cumulative distribution of the load profiles with a month condition 11.5 lies between the distribution of loads in November and December. This demonstrates that the trained CVAE model can generate data using nonexistent conditions during the training process. Also, these profiles have features of data generated using nearby conditions.

STATISTICAL TESTS

To further test the capacity of the CVAE model to generate realistic load profiles, non-visual statistical tests are implemented to inspect different aspects of generations. Specifically, in this experiment, the Kolmogorov-Smirnov test, autoencoder-based test, and energy test were utilized to examine univariate marginal distributions, point-wise multivariate dependencies, and multivariate dependencies of population, respectively. In addition to generated samples with noise $\epsilon \odot \sigma'(\tilde{z}, c)$ added (these were the data used in previous experiments), we also tested the performance of commonly used noise free samples $\mu'(\tilde{z}, c)$.

Evaluating the performance on the Kolmogorov-Smirnov test (Fig. 4.12a), which assesses the accuracy of the marginal distributions, shows a small difference between the training and test sets, and a similar further difference in the distribution accuracy of the generated data. Comparing the results to those reported in [21] for country-level data, we see a slight degradation of the noisy generator. This could be because the individ-

ual load profiles are less smooth than the country-level snapshots, and a relatively large amount of synthetic noise $\epsilon \odot \sigma'(\tilde{z}, c)$ is added to base profiles $\mu'(\tilde{z}, c)$. This can result in the generation of extreme values, which reduces the test scores.

The autoencoder test trains a separate (regular) autoencoder on the training data. This permits the quantification of the quality of individual load profiles. The distributions of reconstruction errors obtained using real and generated data are shown in Fig. 4.12b. The training and test patterns show similar distributions, and the 'noisy' CVAE generates distributions with slightly worse reconstruction errors. However, the 'noise-free' variation produces data that is significantly too smooth, resulting in reconstruction errors that are approximately two orders of magnitude lower.

Finally, the energy test quantifies the similarities between high-dimensional *distributions* of profiles. The results in Fig. 4.12c shows a similar performance between the generated profiles (noisy) and the test data, suggesting good generalization performance. Again, the generated data is a lot more realistic than when no noise is inserted in the output stage ('noise free').

5

CONTROLLABLE GENERATOR: AN ORIENTED VARIATIONAL AUTOENCODER

The functionality and performance of the variational autoencoder-based generators have been evaluated in Chapter 4, using both case studies of country-level load states and individual customers' load profiles. However, not only a generator is required, but generating samples with specific properties is also expected. The link between the latent space codes and generated data may offer the information to infer the properties of the synthetic samples. In this chapter, we propose an oriented variation autoencoder (OVAE) to constrain this link in the form of a Spearman correlation, which provides increased control over the data synthesis process by correlating the features of interest derived from inputs and the data encoded in the latent space. On this basis, an importance sampling process can be used in sampling data in the latent space. Two cases are considered for testing the perfor-

This chapter is based on the following work:

C. Wang, E. Sharifnia, S. Tindemans, and P. Palensky, "Targeted Analysis of High-risk States Using an Oriented Variational Autoencoder", *IEEE Transactions on Power Systems, Submitted.*

The importance sampling and resource adequacy models were developed by Ensieh Sharifnia.

mance of the OVAE model: a) the data set is fully labeled with approximate information; b) the data set is incompletely labeled but with more accurate information. The experimental result shows that, in both cases, the OVAE model makes the data in the latent space to be correlated with the generated data. In addition, the efficiency of generating targeted samples is significantly improved.

5.1. INTRODUCTION

5.1.1. MOTIVATION AND RELATED WORK

The analysis of power system performance across a large range of representative scenarios is of great significance for power system planning and risk assessments [19], [20], [108]. However, data scarcity, reluctance to share, and confidentiality concerns may limit the amount of available historical data, which is a key source of representative scenarios. To this end, it is highly desirable to have access to generative models that produce limitless non-repeating data, reproducing both univariate distributions and interdependencies observed in historical data [78].

In recent years, with the development of deep neural network-related technology, a promising data-driven-based approach has been proposed in the form of *variational autoencoders* (VAEs)[82], [88], [92]. These networks encode high-dimensional historical data (observations) to a latent code in a lower-dimensional *latent space* - and reconstruct similar observations from this code. Features of historical data are learned, so that novel but realistic data points can be created from random codes. In recent research, the VAE model has been successfully used in generating electricity load profiles [23]. Similarly, Generative Adversarial Networks (GANs) have been used to generate realistic power system states [84], [109], but such models do not provide straightforward access to latent representations [110].

Coordinates of latent space codes of VAEs have been shown to correlate with conceptual features of the data [111], [112]. These coordinates can then be used to synthesize targeted data with desired features [113] (e.g., game scenarios). Although the degree of informativeness and orthogonality of latent space variables can be influenced by the training process [95], the interpretation of individual latent variables and the existence of particular concepts are not determined *a priori*.

For power system applications, it is often valuable to generate samples that pertain to certain operating conditions. One use case is when performing studies for a particular time window, geographical area, or otherwise clearly delineated set of conditions. In this case, a generative model can be conditioned on the selection criterion of interest, using e.g. the Conditional VAE (CVAE) instead of the regular VAE. This was done, for example, in [76], for country-level load snapshots conditioned on the hour-of-day and in [114] for 24-hour load profiles of industrial users conditioned on the month-of-year.

A particularly important use case for targeted sampling occurs in power system risk

assessment. As power systems are usually highly reliable, unbiased (Monte Carlo) sampling of states results in excessively high sample count requirements. The accurate estimation of risks can be sped up using importance sampling (IS) [115], [116], which samples high-impact states more often and compensates for the resulting bias by adjusting sample weights. However, this combines the challenge of targeted sampling with (1) knowing *which* states to target and (2) calculating the correct sample weights.

One approach is to use a bottom-up model to generate states and change its parameters for optimal importance sampling. This approach is used, for example, in [117], where generator forced outage rates are modified to speed up generation adequacy assessment, and the cross-entropy method is used to optimize model parameters in a number of stages. This approach can lead to very high speedups, but it requires the availability of a bottom-up generative model and some degree of expert knowledge about which parameters to modify.

Subset simulation represents an alternative approach, where regions of interest in state space are identified and refined in iterations [118]. This requires learning the region of interest and generating states according to this region of interest, ideally without resorting to a sample filtering (accept-reject) approach. A particular challenge is that, to avoid biasing the risk estimate, it is essential that no states with a non-zero impact are excluded from sampling. Moreover, typically no distinction is made between the likelihood of sampling low-impact and high-impact states.

5.1.2. CONTRIBUTION AND OUTLINE

In this chapter, we address the challenges identified above by proposing the *Oriented Variational Autoencoder* (OVAE), a data-driven generative model. Compared to the regular (C)VAE model, it provides increased control over the data synthesis process by correlating the features of interest derived from inputs and the data encoded in the latent space. The model can naturally be used for importance sampling, and can also be trained on incomplete labels. The main contributions of this chapter are as follows:

1) We propose the oriented variational autoencoder (OVAE) that maximizes the Spearman correlation of one latent dimension with a feature of interest.

2) We demonstrate that an OVAE model can be trained in a semi-supervised manner using partially labeled data, which is essential when the process of labels is computationally expensive.

3) Through comprehensive experiments, we test the performance of the OVAE-based generator and its ability to generate calibrated biased samples.

4) The effectiveness of the OVAE model in importance sampling applications is investigated in a case study of multi-area adequacy assessment.

5.2. DATA GENERATION MECHANISM

In this section, a novel multivariate data generation mechanism is proposed, based on the OVAE model. Importance sampling for system adequacy assessment is briefly reviewed in Section 5.2.1. The Oriented VAE and its use in importance sampling are explained in Sections 5.2.2 and 5.2.3, respectively.

5.2.1. IMPORTANCE SAMPLING FOR RISK ASSESSMENT

Quantitative risk assessment for power systems aims to compute one or more numerical risk indices. Often, these are long-run expectations of an operational performance measure, i.e., $r = \mathbb{E}_X[h(X)]$. For example, popular metrics for system adequacy assessment are *Loss Of Load Expectation* (LOLE [h/year]) and *Expected Energy Not Served* (EENS [MWh/year]): LOLE (as measured per hour, also known as LOLH) is the expected number of hours per year during which the supply does not meet demand; EENS is the expected amount of energy demand per year that cannot be supplied. Monte Carlo (MC) simulations can be used to estimate such risks by randomly selecting power system states x (indexed by i) according to their probability density $p(x)$ and calculating the average of the impact $h(x)$ over all sampled states as

$$\hat{r}_{MC} = \frac{1}{m}\sum_{i=1}^{m} h(x_i). \tag{5.1}$$

However, random sampling is computationally inefficient for highly reliable systems when only a small fraction of states contribute, i.e. when $h(x) = 0$ for most states x.

Importance sampling [119] changes the sampling probability distribution to preferentially select samples with higher impact. This reduces the variance of the estimator, and therefore estimates risk values more accurately than the regular Monte Carlo method, for the same number of samples. When states x'_i are sampled according to the modified distribution $q(x)$, the risk is estimated as

$$\hat{r}_{IS} = \frac{1}{m}\sum_{i=1}^{m} h(x'_i)\,w(x'_i), \tag{5.2}$$

with sampling weights $w(x) = \frac{p(x)}{q(x)}$ that ensure an unbiased estimate despite the biased sampling procedure. It is easy to verify that optimal performance is attained when the sampling distribution $q(x)$ is chosen as

$$q^*(x) = \frac{h(x)p(x)}{E_{X\sim p(x)}[h(X)]} = \frac{h(x)p(x)}{r}. \tag{5.3}$$

In fact, this choice reduces the variance to zero, and only a single sample is required. Of course, this distribution is unattainable in practice, as it depends on the quantity-to-be-estimated r in the denominator.

Implementing an effective importance sampling procedure requires the following elements:

- A sufficiently flexible model that generates samples $X \sim q(x;\theta)$, where θ represents parameters that control the sampling distribution.
- Knowledge of the sample impact distribution $h(x)$, so that the generative model can be tuned accordingly.
- An expression for the likelihood ratio (sample weight) $p(x)/q(x)$.

However, in the real world, these requirements are often not met: (1) Common parametric distributions may not be sufficiently flexible to capture complex data distributions. (2) The evaluation of impacts may be computationally expensive. (3) Many generative models do not have an expression for the likelihood ratio, compared to the baseline model. In Section 5.2.3, we will demonstrate that the OVAE model, introduced below, provides a natural framework to address all these challenges.

5.2.2. PROPOSED ORIENTED VAE-BASED GENERATIVE MODEL

For the basic VAE generative model described in Section 4.2, the distribution of latent space codes approximately follows a multivariate standard normal distribution [92], i.e.

$$\frac{1}{n}\sum_{i=1}^{n} q_\phi(z|x_i) \underset{\text{approx.}}{\sim} \mathcal{N}(0,1) \tag{5.4}$$

However, the relation between codes (values of z) and the corresponding states (values of x) or features of interest is otherwise unconstrained. If one aims to perform targeted sampling of states, the location of interesting states in the latent space may not be known, or such states may be distributed in ways that prohibit efficient sampling.

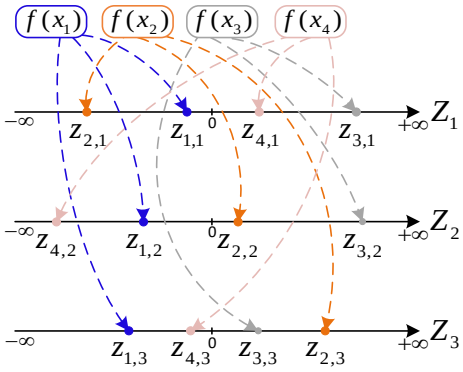

Figure 5.1: Illustration of disordered latent space code.

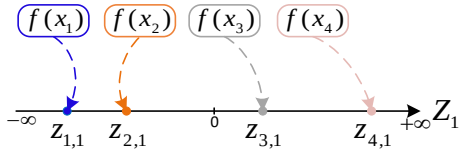

Figure 5.2: Illustration ordered latent space code.

In what follows, $f(x)$ indicates a (single, continuous) feature of interest that can be calculated from the state x. For risk assessment applications, it may be the state impact ($f(x) = h(x)$) or some approximation thereof. In any case, it is assumed to be sufficient to guide targeted sampling of x. Fig. 5.1 illustrates how four different states $(x_1, \ldots, x_4$, in order of increasing $f(x))$ may be mapped onto a three-dimensional latent space. In this case, there is no easily apparent sampling strategy that preferentially targets samples with a particular range of $f(x)$.

Given this, we propose the *oriented variational autoencoder* (OVAE). The idea of OVAE is to force one dimension of the latent space code z to correlate with $f(x)$, while still approximately following a standard normal distribution. In this work, we arbitrarily align increasing values of $f(x)$ with z_1, the first coordinate of the latent space. In the example given above, the aim is to assign z_1 coordinates of samples in order of increasing values $f(x_i)$, as illustrated in Fig. 5.2.

The training process is designed not to disturb the overall distribution of states in the latent space. After training, the known distribution of z_1 and its known alignment with $f(x)$ can be used to perform targeted sampling. The remaining coordinates $z_{l\neq 1}$ can be sampled independently from standard normal distributions as for the regular VAE

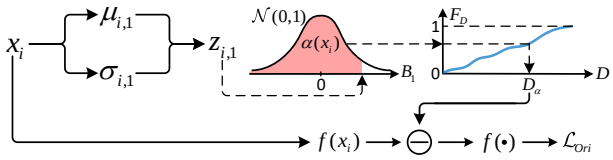

Figure 5.3: Data flow of calculating $\mathscr{L}_{Ori}$.

model. The change in the training process will be achieved by introducing an additional orientation loss term to the training loss (4.7):

$$\mathscr{L} = \beta \mathscr{L}_{D_{KL}} + \mathscr{L}_{Re} + \mathscr{L}_{Ori}. \tag{5.5}$$

The functional form of $\mathscr{L}_{Ori}$ is determined as follows. Conceptually, the desired orientation is achieved by considering the value $f(x)$ as an additional feature that must be reconstructed by the autoencoder network during training, according to the following rules:

- The reconstructed feature (corresponding to a single latent code z) $\hat{f}(z(x_i)) \approx f(x_i)$ is normally distributed, i.e. $\hat{f}(z(x_i)) \sim \mathcal{N}(\mu'_f(z(x_i)), \sigma'_f(z(x_i))$.
- To ensure the desired association of $z_1(x)$ and $f(x)$, only $z_1(x)$ (the first component of the latent code $z(x)$) is used to generate $\mu'_f(z)$. Moreover, this is done using a pre-specified decoder (explained below).
- Unlike other training data, the value $f(x)$ is not used in the encoder, to facilitate semi-supervised learning (explained below).

The decoding function $\mu'_f(z_1)$ is defined as the idealized monotonic mapping from z_1 to $f(x)$. This can be computed *a priori*, by considering that (1) the distribution of z_1 over all samples is a standard normal distribution and (2) using the empirical distribution $\hat{F}(x)$ of $f(x)$ as the target distribution. Using the probability integral transform results in

$$\mu'_f(z_1) = \hat{F}^{-1}(\Phi(z_1)), \tag{5.6}$$

where Φ is the CDF of the standard normal distribution. This procedure is illustrated in Fig. 5.3. Following the steps that led to (4.5) we finally obtain the loss contribution

$$\mathscr{L}_{Ori} = \sum_{i=1}^{n} \left\{ \left[f(x_i) - \hat{F}^{-1}(\Phi(q_\phi(z_{i,1}|x_i))) \right]^2 / \sigma'^2_{f_i} + \log \sigma'^2_{f_i} \right\}. \tag{5.7}$$

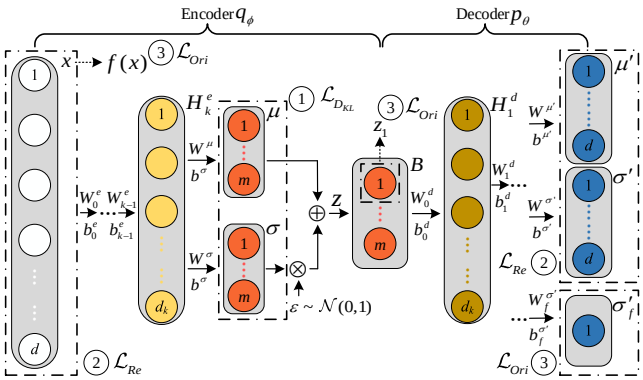

Figure 5.4: Schematic of the proposed OVAE model.

The resulting schematic of the OVAE model is shown in Fig. 5.4. The coordinate z_1 in the latent space is indicated by the dashed box in the bottleneck layer B. Note that, when utilized as a generator, the structure of the OVAE model is the same as that shown in Fig. 4.1b, so the illustration is omitted.

When data labels $f(x)$ are hard-to-get or when the analysis of data labels is time-consuming, data sets may be partially labeled. In this case, the OVAE model can also be trained in a *semi-supervised* manner. When a training batch contains both labeled and unlabeled data, the *Kullback-Leibler loss* $\mathscr{L}_{D_{KL}}$ in (4.4) and *reconstruction loss* $\mathscr{L}_{Re}$ in (4.5) are calculated as an average over all data in the batch, but the orientation loss $\mathscr{L}_{Ori}$ shown in (5.7) is averaged only over the labeled data.

5.2.3. IMPORTANCE SAMPLING WITH OVAE

The OVAE model for data synthesis is highly suitable for use in importance sampling, because it has a known sample distribution in the latent space, and one of the latent space variables encodes a feature of interest. In this work, we use a simple importance sampling procedure that makes use of these properties. The biased sampling distribution in the latent space is given by

$$q(z_1) = \alpha\mathcal{N}(z_1;0,1) + (1-\alpha)\mathcal{N}(z_1;\mu_{IS},\sigma_{IS}^2), \tag{5.8a}$$

$$q(z_{l\neq 1}) = \mathcal{N}(z_l;0,1), \tag{5.8b}$$

where $\mathcal{N}(z;\mu,\sigma^2) = \exp(-(z-\mu)^2/(2\sigma^2))/(\sigma\sqrt{2\pi})$. In this model, the first component z_1 is sampled from a mixture of a standard normal distribution and a normal distribution

with tunable parameters. The other components are sampled from the standard normal distribution. The importance sampling weights are therefore given by

$$w(z) = \frac{p(z)}{q(z)} = \frac{\mathcal{N}(z_1; 0, 1)}{\alpha\mathcal{N}(z_1; 0, 1) + (1-\alpha)\mathcal{N}(z_1; \mu_{IS}, \sigma_{IS}^2)}. \tag{5.9}$$

Exploding weights for relevant states is a known pitfall of importance sampling methods. As can be seen from (5.9), the hyperparameter α ensures that the sampling weight never exceeds $1/\alpha$. A value of 0.1 is used in this work, sacrificing a small fraction of the sample in exchange for additional robustness.

The parameters μ_{IS} and σ_{IS} are chosen such that the importance sampling distribution approximates the optimal sampling distribution (5.3). The procedure for training the OVAE model and the IS parameters for risk assessment (as described in Section 5.2.1) is as follows:

1) Define a feature $f(x)$ that equals or approximates the sample impact $h(x)$.
2) If calculating $f(x)$ is straightforward, label all data points x_i. If not, label a random subset.
3) Train the OVAE model on all data (fully or partially labeled).
4) Use the expectation-maximization (EM) algorithm to optimize μ_{IS} and σ_{IS} so that (5.8a) approximates $q^*(z)$.
5) Sample $\tilde{z}$ according to (5.8), decode latent samples using (4.2a) (inserting output noise in the process) and estimate the risk r using (5.2) and weights (5.9).

5.3. CASE STUDY DESCRIPTION

The OVAE model was trained to generate snapshots of European country-level electricity demand. The generated samples were then used in a multi-area resource adequacy study, where the ability to generate targeted samples was used to greatly increase the sample efficiency. The study case was similar to the one presented in Section 4.4. The data and models used are explained in detail below.

5.3.1. ELECTRICITY DEMAND DATA AND OVAE MODEL STRUCTURE

Historical hourly load demand data for 34 European countries from 2017 and 2018 were obtained from the Open Power System Data platform [48] (package version *2019-06-*

05). The columns of CS (Serbia and Montenegro), IS (Iceland), and UA (Ukraine) were dropped due to incomplete records. Moreover, CY (Cyprus) was omitted due to its lack of connection with power systems from other countries. The historical data were randomly split into training and test sets in blocks of one week with a proportion of 4:1 (13,270 training and 3,212 test samples). The training set was min-max normalized before being fed into the OVAE model, and the inverse transformation was applied to generated samples. The total volume of the generated data was the same as the historical training data set, in order to balance them for visual and statistical analysis. However, we emphasize that the purpose of constructing such a generative model is to have the ability to precisely generate limitless non-repeating data according to users' interests, e.g. to reduce the risk of overfitting in downstream machine learning tasks.

The parameters of the generative models were tuned for optimal performance. The network contained 3 hidden layers in the encoder with dimensions all set as 1,000; the bottleneck layer had 4 nodes (4-dimensional latent vector). The decoder also had 3 hidden layers with the same dimensions as the encoder. The ReLU activation function was used, except for the generation of (μ, σ) and (μ', σ') leading up to the bottleneck and output layers, respectively. The *adaptive moment estimation* (Adam) weight optimizer [51] was utilized with default settings to iteratively optimize the value of weight matrices W and bias vectors b. The batch size was 64, and the learning rate related was 10^{-4}. Training and data generation of the model was conducted in Python using `tensorflow` on the Google Colab environment using the GPU option.

5.3.2. RESOURCE ADEQUACY MODEL

The multi-area resource adequacy model represents the network as a directed graph with flow limits. The topology, capacities, and available generation in each node were based on [103], the 2025 scenario of the ENTSO-E 2020 Mid-term Adequacy Forecast (MAF2020). Net transfer capacities between countries were defined as the summation of transfer capacities between their constituent zones. The released data set does not include generators and unit capacities, so we modeled them as follows. The unit sizes for conventional generators were set on a per-country basis as the closest value under 500 MW that was a divisor of the total capacity of the generators in each country. The assumed unit availability was 80%, and outages were considered independent. The assumed generating capacity of each country was a summation of conventional generating

plus a constant 15% of nameplate wind power capacity. This model is not intended to be an accurate representation of the European network, but to be representative of the studies that can be carried out using the OVAE generative model.

5.3.3. MULTI-AREA RESOURCE ADEQUACY IMPACTS

For a Monte Carlo-based resource adequacy assessment, each sampled state s consists of snapshots of generating capacity $\overline{g}_i$ and demand d_i of each area i. Load curtailment c_i for each node can be calculated based on the flow constraints and dispatching policy using (5.10). This quadratic problem (5.10) with variables $\tilde{c}_i$ (curtailment) and $\tilde{f}_{ij}$ (flows) determines c_i. It minimizes the total load curtailments and finds curtailments balance between areas relative to the demand in that area [104]:

$$c(\overline{g}, d) = \arg\min_{\tilde{f},\tilde{c}} \sum_{i\in\mathcal{N}} \frac{1}{2d_i}\tilde{c}_i^2 + \tilde{c}_i \tag{5.10a}$$

$$\underline{f}_{ij} \le \tilde{f}_{ij} \le \overline{f}_{ij}, \qquad \forall (i\,j) \in \mathcal{L} \tag{5.10b}$$

$$0 \le \tilde{c}_i \le d_i, \qquad \forall i \in \mathcal{N} \tag{5.10c}$$

$$d_i - \overline{g}_i \le \sum_{j<i} \tilde{f}_{ji} - \sum_{j>i} \tilde{f}_{ij} + \tilde{c}_i \le d_i, \qquad \forall i \in \mathcal{N} \tag{5.10d}$$

Here, $\mathcal{L}$ and $\mathcal{N}$ are the sets of connections (from i to j with $i < j$) and areas respectively. Constraints on power flow $\tilde{f}_{ij}$ from node i to node j are given in (5.10b); (5.10c) limits curtailment and (5.10d) enforces flow and generating power constraints. This optimization problem is strictly convex and has a unique solution because the objective function (5.10a) has a positive definite structure, and the constraints are linear. The python package `quadprog` [105] was used to solve this problem.

For a given sampled state $s = (\overline{g}, d)$, the impact $h(\overline{g}, d)$ is calculated using c according to the metric of interest (LOLE or EENS):

$$h_{EENS}(\overline{g}, d) = 8760 \times h_{EPNS}(\overline{g}, d), \tag{5.11}$$

$$h_{LOLE}(\overline{g}, d) = 8760 \times \mathbb{1}_{h_{EPNS}(\overline{g},d)>0}, \tag{5.12}$$

both using the 'power not supplied'

$$h_{EPNS}(\overline{g}, d) = \sum_{i\in\mathcal{N}} c_i(\overline{g}, d). \tag{5.13}$$

Note that by summing curtailments over areas, these are whole-system adequacy metrics, instead of (more common) per-area metrics.

The power system is a very reliable system, so draws from $h_{EPNS}(\overline{g}, d)$ are very likely to return zero. It is undesirable to use a feature with such limited information for orienting the OVAE latent space, so we measure the 'distance from load shedding' for those states where $h_{EPNS}(s) = 0$. This quantity Δ can be defined as the maximum demand that can be added (in proportion to the base demand d) before a shortfall event occurs:

$$\Delta(\overline{g}, d) = \max_{\tilde{f}, \tilde{k}} \tilde{k} \sum_{i \in \mathcal{N}} d_i \tag{5.14a}$$

$$\underline{f}_{ij} \leq \tilde{f}_{ij} \leq \overline{f}_{ij}, \qquad \forall (i\,j) \in \mathcal{L} \tag{5.14b}$$

$$d_i - \overline{g}_i \leq \sum_{j<i} \tilde{f}_{ji} - \sum_{j>i} \tilde{f}_{ij} - \tilde{k} d_i \leq d_i, \qquad \forall\, i \in \mathcal{N} \tag{5.14c}$$

It is easily verified that this has a (non-negative) solution whenever $h_{EPNS}(\overline{g}, d) = 0$.

The auxiliary feature $f(d)$ that is used to train the OVAE model for demand can now be defined in two steps. First, for each demand d, we draw $k = 100$ generation states $\overline{g}^{(j)}$, with $j = 1, \ldots, 100$. Second, we define

$$f_{EENS}(d) = \begin{cases} 8760 \times \overline{h}(d), & \text{if } \overline{h}(d) > 0, \\ -8760 \times \min_j \Delta(\overline{g}^{(j)}, d) & \text{otherwise,} \end{cases} \tag{5.15}$$

where

$$\overline{h}(d) = \frac{1}{100} \sum_{j=1}^{100} h_{EPNS}(\overline{g}^{(j)}, d). \tag{5.16}$$

This is equal to $h_{EENS}(s)$ averaged over 100 generation states when its value is positive, and provides a continuous extension to negative values when no loss of load state is encountered. This makes $f_{EENS}(d)$ suitable as a feature for OVAE alignment.

5.4. EXPERIMENTAL RESULTS OF OVAE

This section describes a number of experiments to test the efficacy of the OVAE model in capturing the data distribution, encoding the feature of interest and potential for importance sampling.

5.4.1. IMPACT OF EXTRA ORIENTED LOSS $\mathscr{L}_{Ori}$ ON MODEL TRAINING

In the first experiment, a simple feature, *total load*, was used to orient the OVAE model. The *total load*

$$f_{TL}(d) = \sum_i d_i \tag{5.17}$$

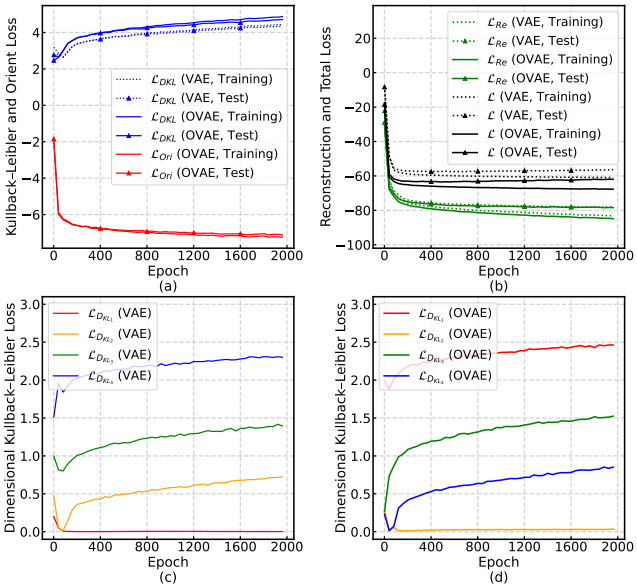

Figure 5.5: Evolution of loss terms during the training of the VAE and OVAE models.

is readily calculated for each load vector d_i and is expected to be correlated with system stress. Like other training data, this feature is normalized. The evolution of the loss terms during the training process of the VAE and OVAE models, for $\beta = 5$, is depicted in Fig. 5.5. The two models show a similar tendency of the *Reconstruction loss* $\mathscr{L}_{Re}$ (Fig. 5.5b). The *Kullback-Leibler loss* $\mathscr{L}_{D_{KL}}$ is slightly higher for the OVAE model, probably due to the additional orientation stress imposed by the *Orientation loss* $\mathscr{L}_{Ori}$ (both Fig. 5.5a). Detailed comparison of the $\mathscr{L}_{D_{KL}}$ loss for each latent dimension (Fig. 5.5c, d) demonstrates that the evolution of the loss (a measure of information contained along its dimension) is similar, but not in identical order. Notably, in the OVAE model, the information content is highest along the first dimension, and for both models, there is an unused dimension (D_{KL}=0). For the OVAE model, the total loss $\mathscr{L}$ of the test set starts increasing slightly after 650 epochs. Thus, to avoid overfitting and to compromise on a good training result, we set the number of training epochs as 650.

5.4.2. VALIDATION OF LATENT SPACE ALIGNMENT

The second experiment investigated the degree of alignment that is achieved between the feature $f_{TL}(d)$ and the latent code $z(d)$. Fig. 5.6a shows a scatter plot of $z_1(d)$ versus

$f_{TL}(d)$ and $z_2(d)$ versus $f_{TL}(d)$, for all training points, using the encoder of the trained OVAE model. The association between z_1 and $f_{TL}(d)$ is clearly visible. To quantify this dependence, the Spearman (rank) correlation coefficient was calculated. This association is maintained for *sampled* data: Fig. 5.6b shows the same for samples $\tilde{z}$ that were generated by sampling from the standard normal distribution in the latent space and the total load $f_{TL}(\tilde{d}(\tilde{z}))$ of the reconstructed snapshot. In contrast, no strong correlation between $f_{TL}(d)$ and $z_1(d)$ (or between $f_{TL}(\tilde{d}(\tilde{z}))$ and $\tilde{z}_1$) is present for the VAE model (Fig. 5.6c, d).

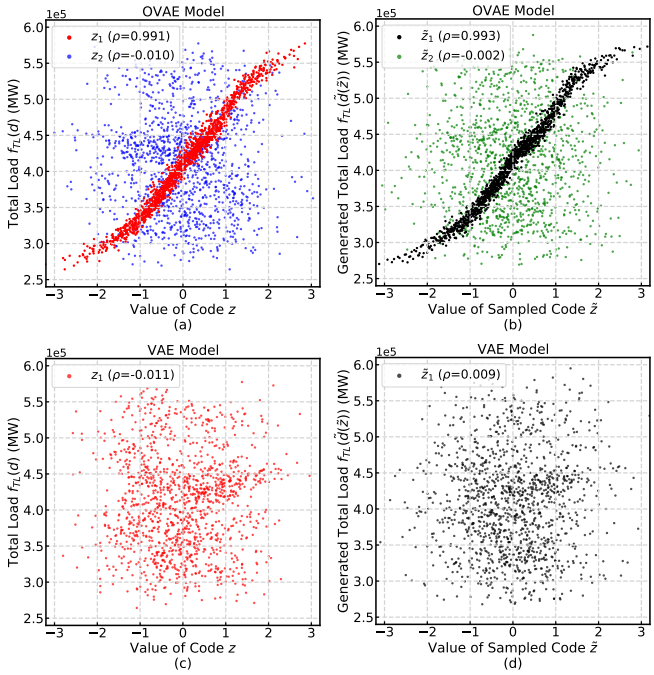

Figure 5.6: Scatter plot and calculated Spearman correlations between total load (input $f_{TL}(d)$ or sampled $f_{TL}(\tilde{d}(\tilde{z}))$) and latent space data (z and $\tilde{z}$) when utilizing OVAE models (a, b) and VAE models (c, d).

5.4.3. UNBIASED AND BIASED SAMPLING

Apart from sampling $\tilde{z}$ from the standard normal distribution $\mathcal{N}(0,1)$, we changed the distribution of $\tilde{z}_1$ and observed the corresponding distribution of the total load $f_{TL}(\tilde{d}(\tilde{z}))$. Note that the other three dimensions of $\tilde{z}$ (i.e., $\tilde{z}_2, \tilde{z}_3$ and $\tilde{z}_4$) were still sampled from the standard normal distribution. Histograms of total load generated by using standard or

nonstandard normal distributed $\tilde{z}_1$ are depicted in Fig. 5.7. When $\tilde{z}_1$ is sampled from $\mathcal{N}(0,1)$, the distribution closely resembles that seen in the historical (training) data. By changing the sampling distribution, targeted generation of low or high load states is possible.

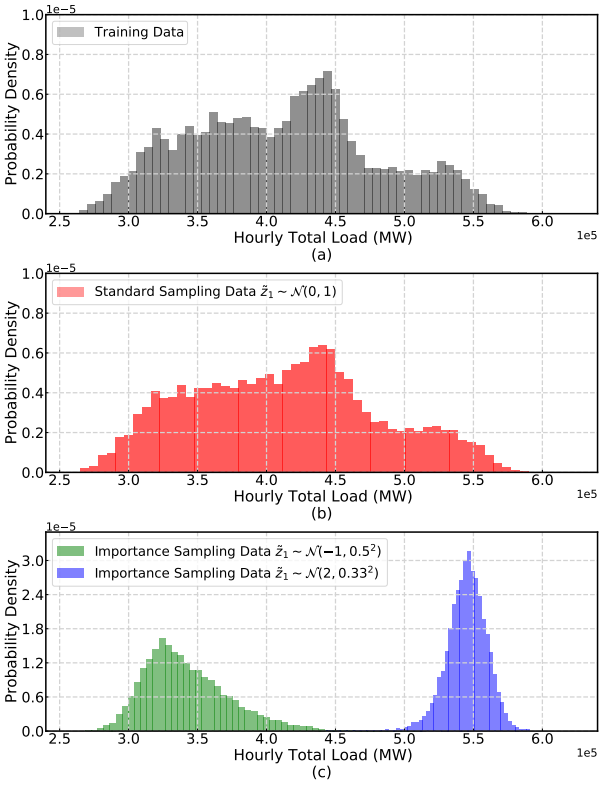

Figure 5.7: Total load of generated samples, compared between training data (top), unbiased sampling from the OVAE model (middle) and biased sampling (two variations) from the OVAE model (bottom).

5.4.4. QUALITY EVALUATION OF GENERATED DATA

To further test the capacity of the CVAE model to generate realistic load profiles, non-visual statistical tests were implemented to inspect different aspects of the generated samples. Specifically, in this experiment, the Kolmogorov-Smirnov test, autoencoder-based test, and energy test were utilized to examine univariate marginal distributions, point-wise multivariate dependencies, and multivariate dependencies of population, respectively (see Section 4.3.2 for a more extensive explanation). The statistical properties

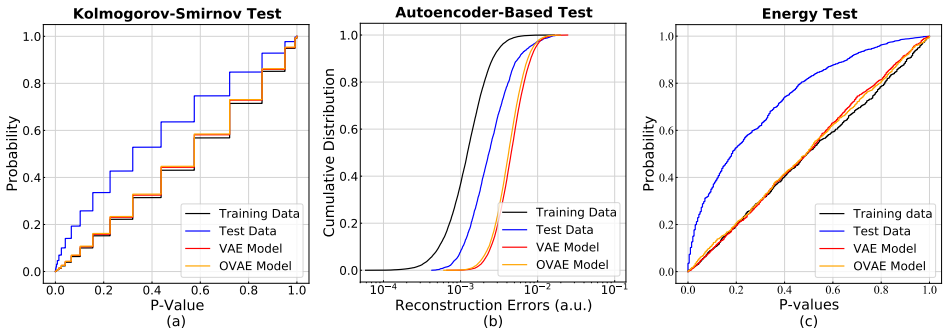

Figure 5.8: Statistical test results of historical data and generated data. (a) Kolmogorov-Smirnov test. (b) Autoencoder-test. (c) Energy test

of four models were studied: the OVAE model, the VAE model, random sampling from the training set, and random sampling from the test set.

The Kolmogorov-Smirnov test (Fig. 5.8a) assesses the accuracy of the marginal distributions. A p-value was calculated by comparing 66 random country-level demand values from the training set (0.5% of the training data set) with the same number of samples from the study model. This was done for each country, and repeated 5,000 times. The results were combined into p-value curves for each study model. The experimental results demonstrate a significant difference between the training and test sets. Compared to the test sets, data generated by the OVAE and VAE models have more similar marginal distributions to that of the training set.

The autoencoder test trains a separate (regular) autoencoder on the training data and tests point-wise multivariate dependencies. The distributions of reconstruction errors obtained using real and generated data are shown in Fig. 5.8b. The results indicate that typical reconstruction errors of demand snapshots generated by VAE and OVAE models are larger than those of the reconstruction errors of training and test distributions. A difference between training and test sets is also visible here.

Finally, the energy test quantifies the similarities of multivariate dependencies of population, compared to the training set. The same as for the K-S test, we used random subsets of 66 data points of the historical and generated population. We used 200 permutations and repeated 1,000 times to draw a distribution of p-values. The results in Fig. 5.8c show that the distribution of samples generated by the OVAE and VAE models is (in this sense) much closer to the training set than that of the test set.

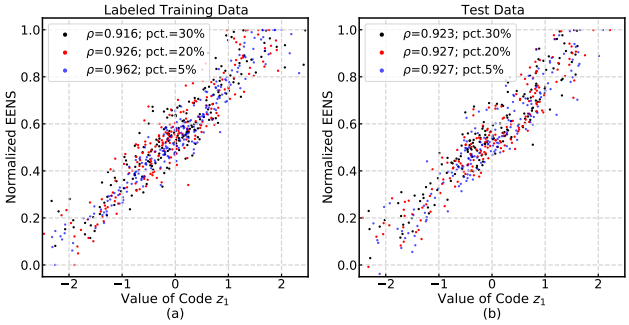

Figure 5.9: Scatter plot and calculated Spearman correlations between the normalized rank of $f_{EENS}(d)$ and latent space data $z_1(d)$ when trained with different percentages of labeled data.

5.4.5. EFFECTIVENESS OF PERFORMING SEMI-SUPERVISED LEARNING

In this section, experiments will be conducted to test if our proposed generator can work properly on incomplete labeled data. Specifically, the OVAE model was trained on a data set that is partially labeled using the computationally intensive label $f_{EENS}(d)$ (5.15) for resource adequacy studies. Different percentages of labeled data were used (5%, 20%, 30%) and labels were replaced by their normalized ranks prior to training.

The experimentally observed dependencies are shown in Fig. 5.9, for training data (left) and test data (right). Using data with different percentages (pct.) of implicit labels for training, $f_{EENS}(d)$ and latent space data z_1 show high Spearman correlations. Notably, data with 5% labels have the highest Spearman correlation, which could be because the small volume of labeled data makes it easier to shape the latent space during training. On the other hand, Spearman correlations are relatively stable on the test data set.

5.4.6. MULTI-AREA ADEQUACY ASSESSMENT RESULTS

Finally, we tested a variety of load models in combination with the resource adequacy model defined in Sections 5.3.2 and 5.3.3. For load models, we considered historical data and 4 OVAE models that were trained with different settings. One used the total load $f_{TL}(d)$ (5.17) as a heuristic feature, and the others were based on the more elaborate $f_{EENS}(d)$ feature ((5.15)). Different percentages of the training data (5%, 20%, and 30%) were labeled.

Simulations to estimate the risk metrics LOLE and EENS were done using 1,000,000

independent samples of demand and available generating capacity. Simulations were implemented in Python 3.8.12 and were run under Windows 10 x64 on a PC equipped with a 4-core Intel Xeon W-2223 CPU (3600 MHz).

To estimate the parameters of the importance sampling distribution (5.8), the procedure described in Section 5.2.3 was followed. The OVAE with standard normal distributions was used to generate 100,000 load states. From this collection of load states, 100,000 states $d^{(j)}$ were drawn at random along with random generating capacity states $\overline{g}^{(j)}$ and the sample weighting function

$$g(\overline{g}^{(j)}, d^{(j)}) = \mathbb{1}_{h_{EPNS}(\overline{g}^{(j)}, d^{(j)})>0} \tag{5.18}$$

was used to assign weights to the sampled points. This selects demand states that cause a shortfall, according to the likelihood for this to happen. Finally, expectation maximization was used to estimate μ_{IS} and σ_{IS} in (5.8a).

Table 5.1 shows the estimated LOLE and EENS values of the European continent for historical load and OVAE models. Risk values are reported in the scientific format, followed by the estimated standard error of the least significant digits in parentheses. For example, $1.190(47) \times 10^4$ stands for an estimate of 11,900 with a standard error of 470. When importance sampling was used, the optimized values for μ_{IS} and σ_{IS} are indicated.

The speedup of sampling-based estimator A with respect to B can be estimated (using the asymptotic speed measure from [120]) as

$$\text{speedup} = \frac{\hat{r}_A^2 t_B \text{SE}(\hat{r}_B)^2}{\hat{r}_B^2 t_A \text{SE}(\hat{r}_A)^2}, \tag{5.19}$$

where t is the execution time of simulation, $\hat{r}$ is the estimated value of the risk metric and $\text{SE}(\hat{r})$ is its standard error. Estimated speedup values for the LOLE and EENS risk metrics are indicated in Table 5.1.

A few conclusions can be drawn from these results. First, all OVAE models, with or without importance sampling, generate LOLE results that are compatible within their margin of error. However, the LOLE results of around 18 hours per year are all higher than that obtained using the historical load values (11 hours per year). That is not unexpected, given the fact that smooth generative models necessarily extrapolate the historical load distribution and will thus generate more extreme demand values.

The gap between historical and generative models increases for the EENS metric that is more sensitive to extreme load values. Here, although the OVAE models trained with

Table 5.1: Resource Adequacy Results and Importance Sampling Speedup

Load model	μ_{IS}	σ_{IS}	Time (s)	LOLE (h/y)	EENS (MWh/y)	LOLE Speedup	EENS Speedup
Historical load	-	-	4319	10.79(31)	$1.190(47) \times 10^4$	n/a	n/a
OVAE-Total Load	0	1	4155	18.76(40)	$4.50(17) \times 10^4$	n/a	n/a
OVAE-EENS 5% training	0	1	4236	18.45(40)	$3.20(10) \times 10^4$	n/a	n/a
OVAE-EENS 20% training	0	1	4130	18.66(40)	$3.46(12) \times 10^4$	n/a	n/a
OVAE-EENS 30% training	0	1	4131	18.16(40)	$2.99(9) \times 10^4$	n/a	n/a
OVAE-Total Load	2.25	0.68	4666	18.43(20)	$4.137(40) \times 10^4$	3.5	14.5
OVAE-EENS 5% training	2.04	0.58	4524	18.06(20)	$3.333(35) \times 10^4$	3.8	8.7
OVAE-EENS 20% training	2.00	0.48	4512	18.55(19)	$3.530(42) \times 10^4$	4.2	7.3
OVAE-EENS 30% training	1.92	0.60	4511	18.17(25)	$3.217(43) \times 10^4$	2.3	5.0

(partial) $f_{EENS}(d)$ labels offer results that are consistent with each other, the OVAE model trained on f_{TL} labels consistently returns higher EENS values. This suggests that the use of the f_{TL} label results in a sample distribution that is slightly heavier in the tails, at least in this instance.

Finally, significant speedups are consistently observed when importance sampling is employed, for both the LOLE and EENS metrics. Although not as large as speedups observed for purpose-made importance sampling schemes [117], it is important to emphasize that the OVAE is a generic data-driven generative model, that can be deployed in a large variety of situations and levels of modeling complexity.

6

CONCLUSION AND RECOMMENDATIONS

This thesis focuses on detecting power system anomalies and generating multivariate load states and profiles. Motivated by the data modeling challenges arising from high dimensional stochastic variables with complex univariate distribution and multivariate dependencies, this thesis proposed an enhanced anomaly detector and novel data generator based on autoencoder neural networks. An autoencoder neural network-based data attack detector was presented, and its hyperparameter selection strategy was investigated. Whitening transformation schemes and whitening matrices selections were investigated to achieve optimal anomaly detection and localization performance, and the performance of anomaly localization was quantified by several novel metrics. In addition, to generate multivariate load data, a conditional variational autoencoder-based data generator has been studied. The impact of output noise and its parameterization on the quality of the generated data was investigated, and the data quality was evaluated in both a visual and statistical manner. Moreover, the performance of the generator was further validated using more stochastic load data from a large variety of individual users. Finally, an oriented variational autoencoder-based data generator has been proposed to synthesize states with a carefully controlled bias. Its performance has been tested with

completely labeled and incompletely labeled training data. Based on the OVAE model, targeted data were sampled with well-defined importance weights.

The following subsections briefly conclude this thesis and provide recommendations for future research.

6.1. CONCLUSIONS

The research questions **Q1** to **Q4** proposed in Chapter 1 have been addressed with both theoretical analysis and experimental validation. The main conclusions of each chapter regarding the research questions are listed as follows.

- **AE-based Anomaly Detection (Q1):** Chapter 2 proposed an FDIA detection approach based on an autoencoder neural network. The proposed detector learns the internal dependency of 'normal' operation data, which avoids the need for gathering attack data for training the classifiers and thus effectively overcomes the inherent unbalanced training data set challenge in the power system. Focusing on 'normal' operating conditions only, novel attacks with features that do not match the patterns inferred from 'normal' data will then be considered anomalies. This one-class classification strategy is well suited for detecting novel attacks that may be fast-evolving, launched by attackers who are resourceful and possibly well-equipped. In addition, the autoencoder can be used as a dimension reduction tool by extracting the lower-dimensional signal from the bottleneck layer.

  Chapter 2 validated the performance of the proposed detector using case studies based on the IEEE 118-bus system: the mechanism is able to robustly detect stealthy FDIAs. Moreover, it still outperforms a BDD scheme when the attacker has only approximate knowledge of the network parameters. In the case study, experiments were conducted to investigate the influence of hyperparameters, i.e., learning rate, batch size, and layer configurations, along with threshold selections on the training process and anomaly detection. The experimental results demonstrate that under proper configurations, the mechanism is able to demonstrate satisfactory learning efficiency and FDIA detection performance. Based on those results, preliminary hyperparameter selection and tuning strategies were put forward.

- **Detector Enhancement (Q2):** Autoencoder neural networks are a powerful tool

for the detection of unknown anomalies. A threshold for the (Euclidean) length of the residuals is typically used to identify anomalous states of a system, but the correlation between residuals is identified as a source of misclassification. Chapter 3 investigated how whitening-based decorrelation of the input features and residuals can improve the performance of the anomaly detector, for a use case of detecting anomalous wind power generation reduction at one or three out of 99 different locations. The reduction of the generated power could represent unexpected malfunction, disturbance, unscheduled outages, or unreported maintenance activities, which are generic anomalies compared to data attacks discussed in Chapter 2. Moreover, the outputs of spatially distributed wind farms are highly variable due to natural variations in wind speed.

Different data processing methods, neural network configuration schemes, and whitening matrix selections were applied to investigate their influence on the performance of autoencoder-based detectors. Whitening of the input data was found to be most beneficial for detection performance across multiple metrics, i.e., ROC, PPV, TNR, ACC, F_1-score. A small further enhancement was obtained when both input data and the residuals were whitened (combined whitening). However, input whitening was found to reduce the ability to locate the source of anomalies. Three metrics, RMS Ratio, Gap Ratio, and OCR, were formulated to quantify this ability. Outstanding localization performance was obtained using standardization of the input data and whitening of the residuals with the ZCA or ZCA-cor whitening matrix. The localization performance was further enhanced by implementing an offset to the whitening transformation.

- **CVAE-based Data Generation (Q3):** Chapter 4 investigated the performance of conditional variational autoencoder- and variational autoencoder-based models to generate multivariate load states. Performance was tested using three statistical tests: Kolmogorov-Smirnov test for univariate marginal distributions, autoencoder-based point-wise test for multivariate dependencies, and energy test for multivariate dependencies of population. A Monte Carlo generation adequacy study on the European network was implemented to illustrate the models' ability to generate realistic tail distribution. In addition to generating snapshots of country-level load states with limited diversity and variability, Chapter 4 also validated the models' capacity to generate synthetic load profiles representing a large variety of individ-

ual users, where the loads are at a lower aggregation level and are more stochastic.

The experimental results demonstrate that the sample noise in the generator and co-optimized output noise parameters lead to generated samples that show better marginal distributions and dependencies when compared with common (C)VAE implementations (fixed noise parameter, noise omitted from the generator). A loss weighting factor β (hyperparameter) can be used to tune the model's performance as a β-VAE [95]. The (C)VAE-based models significantly outperformed Gaussian copula and cGAN models on at least one of the three statistical tests and were competitive on all others. With access to contextual information, the CVAE model slightly outperformed the VAE model. Moreover, such information can be used for target analysis, e.g., as part of a Monte Carlo importance sampling scheme that selects specific hours of the day. The results of generating contextual load profiles of individual customers demonstrate that the proposed model can generate visually realistic profiles that perform well in statistical tests. The results also reconfirm the importance of explicitly including (trained) noise in the final stage of the load profile generator.

- **To achieve a more controllable data generator (Q4):** (C)VAE is demonstrated to generate data with both marginal distributions and multi-variate dependencies embedded. However, naive (C)VAE models don't constrain how latent space codes are placed throughout the standard Gaussian distribution. However, samples with specific properties are valuable for certain applications. Using latent space codes is a promising solution to generate samples with certain properties. To this end, an oriented variational autoencoder (OVAE) was proposed to relate the first dimension of latent space codes and the characteristics of original space data by Spearman correlation. Specifically, apart from the *Kullback-Leibler loss* $\mathscr{L}_{D_{KL}}$ and *reconstruction loss* $\mathscr{L}_{Re}$ terms in (4.5), applied in (C)VAE models, an extra orientation loss $\mathscr{L}_{Ori}$ was used during training to force values of the first dimension codes to increase monotonically with values of the property of interest.

  The performance of the OVAE-based data generator was tested by comprehensive experiments. With an extra orientation loss added, the convergence of the other two losses was affected in a limited way. Moreover, the OVAE generations demonstrated comparable qualities of data generated by the VAE model. When the data set was completely labeled with easy-to-obtain labels (i.e., total load) that are ap-

proximately related to the exact generation goal (i.e., system risk), the experimental results demonstrated not only a good training correlation between data properties and the first dimension codes but also a satisfactory generation correlation between them. Using the trained OVAE model, the generation process of data with user-defined characteristics was significantly accelerated with well-specified importance weight. Additionally, when only a subset (5-30%) of the training data was labeled with precise information on system risk (i.e., EENS), the OVAE model could still be adequately trained, showing an acceptable Spearman correlation between data properties and latent space codes. A case study demonstrated that the OVAE model can be used to speed up power system risk assessment studies.

6.2. DISCUSSION AND RESEARCH RECOMMENDATIONS

The results in this thesis suggest a number of interesting avenues for future research. The detailed recommendations are provided below.

- **Anomaly detection in time-series data:** In Chapter 2 and 3, the (enhanced) autoencoder neural network-based anomaly detectors were trained using only snapshots of system states. However, if the attacker replaces the real system state with a historical state that should not be present at the moment, then the trained detectors are not able to detect it, since no time-series features of the system states have been captured during the training process. In light of this, spatio-temporal measurements can be used as inputs for the autoencoder during training, and the inputs are matrices instead of pure measurement vectors (snapshots). The internal dependencies of each measurement vector, along with the dependencies between those vectors, will be learned together. To do so, the recurrent neural network (RNN) [121] and convolutional neural networks (CNN) [122] are promising solutions to be involved in the detector proposed in Chapter 2 and Chapter 3. Moreover, contextual information, such as the time of day, and day of the week, could be involved in detecting anomalies in time-series data.

- **Anomaly detection in higher-dimensional spaces:** The structure of the autoencoder neural network is naturally suitable for a dimension reduction task of high-dimensional inputs. To validate the anomaly detection performance of the proposed autoencoder-based detectors in high-dimensional space, 339-dimensional

and 99-dimensional measurement vectors were used in Chapter 2 and 3, respectively. However, with the large-scale deployment of sensors, the dimension of measurements will be increasing, and this may degrade the detection sensitivity. The scalability of autoencoder-based detectors for higher-dimensional anomaly detection is worth investigating. To do so, the Polish 2383-bus system [123] may be suitable for a case study. When detecting anomalies in such a high-dimensional space, a whitening transformation, discussed in Chapter 3 for both input and residual data is highly recommended. Moreover, the partition of a large-scale power grid and considering only data measured in small parts could be another promising solution to address the challenges of extremely high-dimensional data.

- **Automatic hyperparameter selection for anomaly detector:** In Chapter 2 and 3, the hyperparameters of the autoencoder neural networks were set using a trial-and-error process. This manual process is rather time-consuming because of the large number of hyperparameters, e.g., number of layers, dimension of layers, learning rate, batch size, activation function, training epochs, initialization of the weight and bias, and optimization algorithm. The combinations of those hyperparameters are numerous, and there could be a long training process under each combination. In addition, the hyperparameter tuning process also requires specialist experience and rules of thumb. In view of this, investigating automated and computationally efficient hyperparameter tuning strategies is appealing. Random search [124], and Bayesian Optimization [125] may be possible solutions to find satisfactory hyperparameter values with fewer function evaluations [126].

- **More expressivity of the decoder:** In Chapter 4, based on the (conditional) variational autoencoder, data generators were proposed to synthesize country-level load states and individual load profiles. These generators were trained by a constraint that the output noise on each dimension is independent. Due to the complex multi-dimensional distribution involved in real-world data, generating data with noise that is dimensionally correlated may help to synthesize data with a more realistic distribution. Cholesky decomposition could be a solution to make the noise distributed with full covariance. The equation shown in (4.2b) can be changed to $\hat{x} = \mu'(\hat{z}) + L(\hat{z}) \cdot \epsilon$, where L is the lower triangular matrix of the output noise. Notably, the quality of the generated data is also required to be verified using the metrics used in Chapter 4. Initial results show that the additional flexibility

of correlated output noise frequently results in unstable training.

- **Embedding of physical constraints in data generators:** In Chapter 4, the generated load states and load profiles were decoded from codes sampled from multi-dimensional standard Gaussian distributions. However, the generation process is not constrained by any physical limits, thus there is a probability that the generated load data exceeds the upper or lower limits of possible loads. A brute-force truncated Gaussian distribution might restrain the trained generator from synthesizing extreme values. However, this could, in turn, affect the quality of the generated data, such as the marginal distribution. In light of this, curiosity is aroused about how to embed physical constraints during the training process. A possible solution could be adding an extra loss term to the loss function as a Physics-informed neural network (PINN) [127]. This extra loss is related to the value of $\hat{x}$ shown in (4.2b). Once the value of $\hat{x}$ exceeds the set constraints, a large penalty will be imposed on the loss value. However, the implementation details and mathematical derivations need to be carefully considered.

BIBLIOGRAPHY

[1] U. E. I. Administration, *Electricity consumption*, 2022. Accessed on: Nov. 2, 2022. [Online]. Available: https://www.eia.gov/international/data/world/.

[2] U. E. I. Administration, *Electricity capacity*, 2022. Accessed on: Nov. 2, 2022. [Online]. Available: https://www.eia.gov/international/data/world/.

[3] S. E. International, *Asia to deploy more than 570 million smart electricity meters*, 2021. Accessed on: Nov. 2, 2022. [Online]. Available: https://www.smart-energy.com/industry-sectors/smart-meters/smart-electricity-meters-rollout-in-china-india-japan-and-south-korea/.

[4] M. Sun, I. Konstantelos, S. Tindemans, and G. Strbac, "Evaluating composite approaches to modelling high-dimensional stochastic variables in power systems", in *2016 Power Systems Computation Conference (PSCC)*, IEEE, 2016, pp. 1–8.

[5] O. Ardakanian, S. Keshav, and C. Rosenberg, "Real-time distributed control for smart electric vehicle chargers: From a static to a dynamic study", *IEEE Transactions on Smart Grid*, vol. 5, no. 5, pp. 2295–2305, 2014.

[6] S. Tindemans, P. Djapic, J. Schofield, T. Ustinova, and G. Strbac, "Resilience performance of smart distribution networks: Report d4 for the "low carbon london" lcnf project", 2017.

[7] R. Liu, C. Vellaithurai, S. S. Biswas, T. T. Gamage, and A. K. Srivastava, "Analyzing the cyber-physical impact of cyber events on the power grid", *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2444–2453, 2015.

[8] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price", *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 627–636, 2013.

[9] A. Venzke, D. K. Molzahn, and S. Chatzivasileiadis, "Efficient creation of datasets for data-driven power system applications", *Electric Power Systems Research*, vol. 190, p. 106 614, 2021.

[10] F. Thams, A. Venzke, R. Eriksson, and S. Chatzivasileiadis, "Efficient database generation for data-driven security assessment of power systems", *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 30–41, 2019.

[11] S. Kumar, "Classification and detection of computer intrusions", Ph.D. dissertation, Purdue University, 1995.

[12] D. M. Hawkins, *Identification of outliers*. Springer, 1980, vol. 11.

[13] C. C. Aggarwal, "An introduction to outlier analysis", in *Outlier analysis*, Springer, 2017, pp. 1–34.

[14] T. M. Chen and S. Abu-Nimeh, "Lessons from stuxnet", *Computer*, vol. 44, no. 4, pp. 91–93, 2011.

[15] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 ukraine blackout: Implications for false data injection attacks", *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317–3318, 2016.

[16] D. U. Case, "Analysis of the cyber attack on the ukrainian power grid", *Electricity Information Sharing and Analysis Center (E-ISAC)*, vol. 388, pp. 1–29, 2016.

[17] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks", *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1244–1253, 2013.

[18] W. Li, *Risk assessment of power systems: models, methods, and applications*. John Wiley & Sons, 2014.

[19] P. Panciatici, G. Bareux, and L. Wehenkel, "Operating in the fog: Security management under uncertainty", *IEEE Power and Energy Magazine*, vol. 10, no. 5, pp. 40–49, 2012.

[20] H. Bloomfield, D. Brayshaw, A. Troccoli, *et al.*, "Quantifying the sensitivity of European power systems to energy scenarios and climate change projections", *Renewable Energy*, vol. 164, pp. 1062–1075, 2021.

[21] C. Wang, E. Sharifnia, Z. Gao, S. H. Tindemans, and P. Palensky, "Generating multivariate load states using a conditional variational autoencoder", *arXiv preprint arXiv:2110.11435*, 2021.

[22] X. Gong, B. Tang, R. Zhu, W. Liao, and L. Song, "Data augmentation for electricity theft detection using conditional variational auto-encoder", *Energies*, vol. 13, no. 17, p. 4291, 2020.

[23] Z. Pan, J. Wang, W. Liao, *et al.*, "Data-driven EV load profiles generation using a variational auto-encoder", *Energies*, vol. 12, no. 5, p. 849, 2019.

[24] C. Wang, S. Tindemans, K. Pan, and P. Palensky, "Detection of false data injection attacks using the autoencoder approach", in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, IEEE, 2020, pp. 1–6.

[25] C. C. Aggarwal, "Outlier analysis", in *Data mining*, Springer, 2015, pp. 237–263.

[26] K. Yamanishi and J.-i. Takeuchi, "Discovering outlier filtering rules from unlabeled data: Combining a supervised learner with an unsupervised learner", in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 389–394.

[27] E. Eskin, "Anomaly detection over noisy data using learned probability distributions", 2000.

[28] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "Knn model-based approach in classification", in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, Springer, 2003, pp. 986–996.

[29] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers", in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[30] C. Wang, K. Pan, S. Tindemans, and P. Palensky, "Training strategies for autoencoder-based detection of false data injection attacks", in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, 2020, pp. 1–5.

[31] H. Liu, F. Han, M. Yuan, J. Lafferty, and L. Wasserman, "High-dimensional semiparametric Gaussian copula graphical models", *The Annals of Statistics*, vol. 40, no. 4, pp. 2293–2326, 2012.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial networks", *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[33] E. Wilton, E. Delarue, W. D'haeseleer, and W. van Sark, "Reconsidering the capacity credit of wind power: Application of cumulative prospect theory", *Renewable energy*, vol. 68, pp. 752–760, 2014.

[34] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids", *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[35] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm", *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[36] J. Shlens, "A tutorial on principal component analysis", *arXiv preprint arXiv:1404.1100*, 2014.

[37] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction", in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, ACM, 2014, p. 4.

[38] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using kalman filter", *IEEE transactions on control of network systems*, vol. 1, no. 4, pp. 370–379, 2014.

[39] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids", *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, 2015.

[40] J. Zhao, L. Mili, and M. Wang, "A generalized false data injection attacks against power system nonlinear state estimator and countermeasures", *IEEE Transactions on Power Systems*, p. 1, 2018, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2018.2794468.

[41] K. Pan, P. Palensky, and P. M. Esfahani, "From static to dynamic anomaly detection with application to power system cyber security", *IEEE Transactions on Power Systems*, pp. 1–1, 2019. DOI: 10.1109/tpwrs.2019.2943304.

[42] J. Tian, M. H. Azarian, and M. Pecht, "Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm", in *Proceedings of the European Conference of the Prognostics and Health Management Society*, Citeseer, 2014.

[43] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism", *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, 2017.

[44] J. James, Y. Hou, and V. O. Li, "Online false data injection attack detection with wavelet transform and deep neural networks", *IEEE Transactions on Industrial Informatics*,

[45] F. M. Gonzalez-Longatt and J. L. Rueda, *PowerFactory applications for power system analysis*. Springer, 2014.

[46] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks", in *First Workshop on Secure Control Systems (SCS), Stockholm*, 2010.

[47] L. Duan, M. Xie, T. Bai, and J. Wang, "A new support vector data description method for machinery fault diagnosis with unbalanced datasets", *Expert Systems with Applications*, vol. 64, pp. 239–246, 2016.

[48] J. Muehlenpfordt, "Time series", *Open Power System Data*, 2019. [Online]. Available: https://data.open-power-system-data.org/time%20series/2019-06-05.

[49] R. D. Zimmerman, C. E. Murillo-Sánchez, and D. Gan, "Matpower: A matlab power system simulation package", *Manual, Power Systems Engineering Research Center, Ithaca NY*, vol. 1, 1997.

[50] M. He, V. Vittal, and J. Zhang, "Online dynamic security assessment with missing pmu measurements: A data mining approach", *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1969–1977, 2013.

[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

[52] C. Wang, K. Pan, S. Tindemans, and P. Palensky, "Training strategies for autoencoder-based detection of false data injection attacks", in *The 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe 2020)*, arXiv:2005.07158, 2020.

[53] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, "Cyber risk analysis of combined data attacks against power system state estimation", *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3044–3056, 2018.

[54] Y. Wang, N. Zhang, C. Kang, M. Miao, R. Shi, and Q. Xia, "An efficient approach to power system uncertainty analysis with high-dimensional dependencies", *IEEE Transactions on Power Systems*, pp. 2984–2994, 2017.

[55] H. Zhao, H. Liu, W. Hu, and X. Yan, "Anomaly detection and fault analysis of wind turbine components based on deep learning network", *Renewable energy*, vol. 127, pp. 825–834, 2018.

[56] L. Wang, Z. Zhang, J. Xu, and R. Liu, "Wind turbine blade breakage monitoring with deep autoencoders", *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2824–2833, 2016.

[57] Y. Li, W. Jiang, G. Zhang, and L. Shu, "Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data", *Renewable Energy*, vol. 171, pp. 103–115, 2021.

[58] E. Vladislavleva, T. Friedrich, F. Neumann, and M. Wagner, "Predicting the energy output of wind farms based on weather data: Important variables and their correlation", *Renewable energy*, vol. 50, pp. 236–243, 2013.

[59] T. Denouden, R. Salay, K. Czarnecki, V. Abdelzad, B. Phan, and S. Vernekar, "Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance", *arXiv preprint arXiv:1812.02765*, 2018.

[60] G. Jiang, P. Xie, H. He, and J. Yan, "Wind turbine fault detection using a denoising autoencoder with temporal information", *IEEE/Asme transactions on mechatronics*, vol. 23, no. 1, pp. 89–100, 2017.

[61] J. Renman, "Deep autoencoder for condition monitoring of wind turbines-detecting and diagnosing anomalies", M.S. thesis, Chalmers Tekniska Högskola, Gothenburg, Sweden, 2019.

[62] Y. Cui, P. Bangalore, and L. B. Tjernberg, "An anomaly detection approach based on machine learning and scada data for condition monitoring of wind turbines", in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, IEEE, 2018, pp. 1–6.

[63] A. Kessy, A. Lewin, and K. Strimmer, "Optimal whitening and decorrelation", *The American Statistician*, vol. 72, no. 4, pp. 309–314, 2018.

[64] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning cnn image retrieval with no human annotation", *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1655–1668, 2018.

[65] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos", *Journal of Imaging*, vol. 4, no. 2, p. 36, 2018.

[66] N. Renström, P. Bangalore, and E. Highcock, "System-wide anomaly detection in wind turbines using deep autoencoders", *Renewable Energy*, vol. 157, pp. 647–659, 2020.

[67] G. Li and J. Zhang, "Sphering and its properties", *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 119–133, 1998.

[68] J. H. Friedman, "Exploratory projection pursuit", *Journal of the American statistical association*, vol. 82, no. 397, pp. 249–266, 1987.

[69] P. C. Mahalanobis, "On the generalised distance in statistics", *Proceedings of the National Institute of Sciences of India*, vol. 2, pp. 49–55, 1936.

[70] Wikipedia, "North holland", *wikipedia*, 2010. [Online]. Available: https://en.wikipedia.org/wiki/North_Holland.

[71] Wikipedia, "List of municipalities in south holland", *wikipedia*, 2010. [Online]. Available: https://en.wikipedia.org/wiki/List_of_municipalities_in_South_Holland.

[72] S. Pfenninger and I. Staffell, "Renewables.ninja", *Renewables.ninja*, 2017. [Online]. Available: https://www.renewables.ninja/.

[73] M. M. Rienecker, M. J. Suarez, R. Gelaro, R. Todling, *et al.*, "Merra: Nasa's modern-era retrospective analysis for research and applications", *Journal of climate*, vol. 24, no. 14, pp. 3624–3648, 2011.

[74] I. Staffell and S. Pfenninger, "Using bias-corrected reanalysis to simulate current and future wind power output", *Energy*, vol. 114, pp. 1224–1239, 2016.

[75] T. Fawcett, "An introduction to roc analysis", *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[76] C. Wang, E. Sharifnia, Z. Gao, S. H. Tindemans, and P. Palensky, "Generating multivariate load states using a conditional variational autoencoder", *Electric Power Systems Research*, vol. 213, p. 108 603, 2022.

[77] C. Wang, S. H. Tindemans, and P. Palensky, "Generating contextual load profiles using a conditional variational autoencoder", in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, IEEE, 2022, pp. 1–6.

[78] I. Konstantelos, M. Sun, S. H. Tindemans, S. Issad, P. Panciatici, and G. Strbac, "Using vine copulas to generate representative system states for machine learning", *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 225–235, 2019. DOI: 10.1109/TPWRS.2018.2859367.

[79] M.-S. Kang, C.-S. Chen, Y.-L. Ke, C.-H. Lin, and C.-W. Huang, "Load profile synthesis and wind-power-generation prediction for an isolated power system", *IEEE Transactions on Industry Applications*, vol. 43, no. 6, pp. 1459–1464, 2007.

[80] F. B. dos Reis, R. Tonkoski, and T. M. Hansen, "Synthetic residential load models for smart city energy management simulations", *IET Smart Grid*, vol. 3, no. 3, pp. 342–354, 2020.

[81] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, 2nd ed. Academic Press, 2020, pp. 67–120, ISBN:978-0-12-818803-3, ISBN: 978-0-12-818803-3.

[82] D. P. Kingma and M. Welling, "Auto-encoding variational bayes", *arXiv preprint arXiv:1312.6114*, 2013.

[83] Y. Gu, Q. Chen, K. Liu, L. Xie, and C. Kang, "Gan-based model for residential load generation considering typical consumption patterns", in *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, IEEE, 2019, pp. 1–5.

[84] Z. Wang and T. Hong, "Generating realistic building electrical load profiles through the generative adversarial network (gan)", *Energy and Buildings*, vol. 224, p. 110 299, 2020.

[85] R. Yu, "A tutorial on VAEs: From Bayes' rule to lossless compression", *arXiv preprint arXiv:2006.10273*, 2020.

[86] S. Lin, S. Roberts, N. Trigoni, and R. Clark, "Balancing reconstruction quality and regularisation in elbo for vaes", *arXiv preprint arXiv:1909.03765*, 2019.

[87] O. Rybkin, K. Daniilidis, and S. Levine, "Simple and effective VAE training with calibrated decoders", in *International Conference on Machine Learning*, PMLR, 2021, pp. 9179–9189.

[88] C. Doersch, "Tutorial on variational autoencoders", *arXiv preprint arXiv:1606.05908*, 2016.

[89] C. Mylonas, I. Abdallah, and E. Chatzi, "Conditional variational autoencoders for probabilistic wind turbine blade fatigue estimation using supervisory, control, and data acquisition data", *Wind Energy*, vol. 24, pp. 1122–1139, 10 2021.

[90] Y. Qi, W. Hu, Y. Dong, Y. Fan, L. Dong, and M. Xiao, "Optimal configuration of concentrating solar power in multienergy power systems with an improved variational autoencoder", *Applied Energy*, vol. 274, p. 115 124, 2020.

[91] M. Brégère and R. J. Bessa, "Simulating tariff impact in electrical energy consumption profiles with conditional variational autoencoders", *IEEE Access*, vol. 8, pp. 131 949–131 966, 2020.

[92] D. P. Kingma, M. Welling, *et al.*, "An introduction to variational autoencoders", *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[93] J. Xu and G. Durrett, "Spherical latent spaces for stable variational autoencoders", *arXiv preprint arXiv:1808.10805*, 2018.

[94] W. Joo, W. Lee, S. Park, and I.-C. Moon, "Dirichlet variational autoencoder", *Pattern Recognition*, vol. 107, p. 107 514, 2020.

[95] C. P. Burgess, I. Higgins, A. Pal, *et al.*, "Understanding disentangling in β-VAE", *arXiv preprint arXiv:1804.03599*, 2018.

[96] C. Wang, *Code release: Generating multivariate load states using a conditional variational autoencoder*, 2022. Accessed on: Apr. 13, 2022. [Online]. Available: https://github.com/ChenguangWang-Sam/PSCC2022-CVAE.

[97] E. Sharifnia, *Code release: System adequacy case study for CVAE load generation, PSCC 2022*, 2022. Accessed on: Apr. 15, 2022. [Online]. Available: https://github.com/ensieh-sharifnia/MC-PSCC2022.

[98] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit", *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[99] G. J. Székely and M. L. Rizzo, "Energy statistics: A class of statistics based on distances", *Journal of statistical planning and inference*, vol. 143, no. 8, pp. 1249–1272, 2013.

[100] J. Djolonga, *A PyTorch library for differentiable two-sample tests*, 2017. Accessed on: Oct. 3, 2021. [Online]. Available: https://github.com/josipd/torch-two-sample/blob/master/docs/index.rst.

[101] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial nets", *Advances in neural information processing systems*, vol. 27, 2014.

[102] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks", in *International conference on machine learning*, PMLR, 2017, pp. 214–223.

[103] ENTSO-E, *Mid-term adequacy forecast 2020*, 2020. Accessed on: Oct. 1, 2021. [Online]. Available: https://www.entsoe.eu/outlooks/midterm/.

[104] M. P. Evans and S. H. Tindemans, "Assessing energy storage requirements based on accepted risks", in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, 2020, pp. 1109–1113.

[105] R. T. McGibbon, *Quadprog 0.1.8 - pypi*, 2021. Accessed on: Aug. 15, 2021. [Online]. Available: https://pypi.org/project/quadprog.

[106] nl.wikipedia.org, "Alliander", *Wikipedia*, 2022. Accessed on: May. 16, 2022. [Online]. Available: https://nl.wikipedia.org/wiki/Alliander.

[107] S. Lloyd, "Least squares quantization in pcm", *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[108] H. Wang, Y.-P. Fang, and E. Zio, "Risk assessment of an electrical power system considering the influence of traffic congestion on a hypothetical scenario of electrified transportation system in new york state", *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 142–155, 2019.

[109] G. Baasch, G. Rousseau, and R. Evins, "A conditional generative adversarial network for energy use in multiple buildings using scarce data", *Energy and AI*, vol. 5, p. 100 087, 2021.

[110] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview", *IEEE signal processing magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[111] T. Spinner, J. Körner, J. Görtler, and O. Deussen, "Towards an interpretable latent space: An intuitive comparison of autoencoders with variational autoencoders", in *IEEE VIS 2018*, 2018.

[112] G. P. Way and C. S. Greene, "Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders", in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium*, World Scientific, 2018, pp. 80–91.

[113] A. Sarkar and S. Cooper, "Generating and blending game levels via quality-diversity in the latent space of a variational autoencoder", in *The 16th International Conference on the Foundations of Digital Games (FDG) 2021*, 2021, pp. 1–11.

[114] C. Wang, S. H. Tindemans, and P. Palensky, "Generating contextual load profiles using a conditional variational autoencoder", *arXiv preprint arXiv:2209.04056*, 2022.

[115] Q. Chen and L. Mili, "Composite power system vulnerability evaluation to cascading failures using importance sampling and antithetic variates", *IEEE transactions on power systems*, vol. 28, no. 3, pp. 2321–2330, 2013.

[116] A. B. Owen, Y. Maximov, and M. Chertkov, "Importance sampling the union of rare events with an application to power systems analysis", *Electronic Journal of Statistics*, vol. 13, no. 1, pp. 231–254, 2019.

[117] A. M. L. Da Silva, R. A. Fernandez, and C. Singh, "Generating capacity reliability evaluation based on monte carlo simulation and cross-entropy methods", *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 129–137, 2010.

[118] Y. Zhao, Y. Han, Y. Liu, K. Xie, W. Li, and J. Yu, "Cross-entropy-based composite system reliability evaluation using subset simulation and minimum computational burden criterion", *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5198–5209, 2021.

[119] B. Roy and L. Wenyuan, *Reliability assessment of electric power systems using Monte Carlo methods*. Springer Science & Business Media, 2013.

[120] S. Tindemans and G. Strbac, "Accelerating system adequacy assessment using the multilevel monte carlo approach", *Electric Power Systems Research*, vol. 189, p. 106 740, 2020.

[121] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization", *arXiv preprint arXiv:1409.2329*, 2014.

[122] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network", in *2017 international conference on engineering and technology (ICET)*, Ieee, 2017, pp. 1–6.

[123] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education", *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2010.

[124] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.", *Journal of machine learning research*, vol. 13, no. 2, 2012.

[125] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms", *Advances in neural information processing systems*, vol. 25, 2012.

[126] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.

[127] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning", *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

CURRICULUM VITÆ

Chenguang Wang was born on April 23, 1992, in Jingzhou, Hubei Province, China. He obtained his B.Sc. degree in Electrical Engineering from Wuhan University of Technology in 2014, where he devoted much of his spare time to researching power electronics technologies and conducting experiments in the laboratory. During his undergraduate studies, he won the first prize in the "National Undergraduate Electronics Design Contest (Hubei Division)," which is one of the most prestigious student competitions in electrical engineering. With achievements in student contests and high academic grades, he was recommended as an exam-exempted postgraduate to Xi'an Jiaotong University, where he pursued a master's degree in High Voltage and Insulation Technology in Electrical Engineering.

In 2017, Chenguang graduated from Xi'an Jiaotong University with an Outstanding Graduate award, having received several awards and scholarships during his master's studies, including the 2014-2015 school year "China National scholarship" and the grand prize of the "8^{th} University Student Social Practice and Science Contest on Energy Saving & Emission Reduction," which is a highly regarded academic competition in China. With only 0.36% of participants receiving the grand prize, Chenguang led his team to rank 9^{th} out of over 2500 teams.

In 2018, Chenguang joined the Intelligent Electrical Power Grids group at Delft University of Technology to pursue his Ph.D. degree under the supervision of Prof.dr. Peter Palensky and Dr. Simon Tindemans. During his Ph.D. journey, he focused on developing state-of-the-art machine learning models to address the challenges of power system anomaly detection and synthetic data generation. As the first author, he published and submitted 6 high-level research papers in top conferences and journals, attracting attention from both industry and academia due to their practicality.

List of Publications

C. Wang, S. Tindemans, K. Pan, and P. Palensky, "Detection of false data injection attacks using the autoencoder approach", in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, IEEE, Liege, Belgium, 2020, pp. 1–6. DOI: 10.1109/PMAPS47429.2020.9183526

C. Wang, K. Pan, S. Tindemans, and P. Palensky, "Training strategies for autoencoder-based detection of false data injection attacks", in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, Den Haag, the Netherlands, 2020, pp. 1–5. DOI: 10.1109/ISGT-Europe47291.2020.9248894

C. Wang, E. Sharifnia, Z. Gao, S. H. Tindemans, and P. Palensky, "Generating multivariate load states using a conditional variational autoencoder", *presented* in *XXII Power Systems Computation Conference (PSCC 2022)*, Porto, Portugal, 2022 and *published* in *Electric Power Systems Research*, vol. 213, p. 108603, 2022.

C. Wang, S. Tindemans, and P. Palensky, "Improved Anomaly Detection and Localization Using Whitening-Enhanced Autoencoders", *IEEE Transactions on Industrial Informatics, Accepted.* DOI: 10.1109/TII.2023.3268685

C. Wang, S. H. Tindemans, and P. Palensky, "Generating contextual load profiles using a conditional variational autoencoder", in *2022 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, IEEE, Novi Sad, Serbia, 2022, pp. 1–6. DOI: 10.1109/ISGT-Europe54678.2022.9960309

C. Wang, E. Sharifnia, S. Tindemans, and P. Palensky, "Targeted Analysis of High-risk States Using an Oriented Variational Autoencoder", *IEEE Transactions on Power System, Submitted.*

ACKNOWLEDGEMENTS

Throughout my Ph.D. journey, I have been fortunate to have the support and guidance of many individuals who generously shared their time, expertise, and resources. I am deeply grateful for the invaluable insights, feedback, and encouragement that they provided. I would like to express my heartfelt gratitude to all those who have made this Ph.D. thesis possible.

First of all, I would like to express my deep gratitude to Prof.dr. P. Palensky, my Ph.D. promoter, for his exceptional support, guidance, and expertise throughout my doctoral studies. Peter's efficiency, excellent communication skills, and commitment to my research have been vital in the successful completion of my thesis. I have learned so much from Peter's vast knowledge and experience, as he always kept my research on the right track and provided invaluable feedback. He also shared relevant conference and journal paper opportunities with me promptly, ensuring that I was constantly updated on the latest developments in my field.

Moreover, Peter not only cared about my academic growth but also my mental well-being during the COVID-19 pandemic. When I was stranded outside the Netherlands due to travel restrictions, he frequently reached out to me and offered encouragement and support. His emails and messages helped me stay motivated and focused on my research and personal development. I feel incredibly fortunate to have had Peter as my promoter, and I will always be grateful for his generous and steadfast mentorship.

Next, I would like to express my sincere gratitude to Dr. S.H. Tindemans, my copromotor, for his invaluable guidance throughout my Ph.D. journey. Simon's dedication and expertise have been instrumental in shaping my research and improving my academic writing. I am deeply grateful to him for revising my academic papers late into the night and even taking time out of his vacation to do so. Simon not only helped me with formatting, symbols, and expressions, but also emphasized the importance of precision in describing my research data and provided insightful suggestions to refine my scientific experiments.

In addition, Simon has inspired me with new research ideas, which have broadened

my perspective and challenged me to think more critically. Our weekly meetings have been the highlight of my research experience. Simon's thought-provoking questions have often made my "head spin", but they have always encouraged me to make further progress. His excellent advice has helped me avoid time-consuming pitfalls and to stay focused on what matters most. Most importantly, Simon has taught me the value of "baby steps" in research. He has always motivated me to take small, achievable steps toward my goals, rather than trying to tackle everything at once. This approach has been a game-changer for me, and I am grateful for Simon's wisdom and direction. Thank you, Simon, for being an outstanding copromotor. Your support and inspiration have made all the difference, and I am honored to have had you as my copromotor.

I would like to convey my deep appreciation to my parents for their consistent and unwavering support over the past 30 years. Particularly, I want to thank my father for collecting every achievement I made and compiling them into a special album. This album not only reflects my academic growth, but also serves as a constant reminder of the immense love and pride my father has for me. His steadfast dedication and support have been a constant source of motivation for me throughout my academic pursuits. It is my father's emphasis on education that has brought me to where I am today. I would also like to express my sincere appreciation to my mother for her unwavering love and support. Her infectious passion for life, including her love of singing and travel, and her enthusiasm for sharing her life highlights have consistently inspired me to see the beauty in everyday moments. The photos and videos she shared with me have filled me with joy and provided me with a sense of relaxation during my studies. Her positive attitude and loving care, evident in her frequent phone calls, have helped me to face the challenges of life with optimism and perseverance. I am deeply grateful to have them as my parents and to know that I can always count on their support in every aspect of my life. I would also like to extend my heartfelt gratitude to my parents-in-law and family for their unconditional support during my Ph.D. studies abroad. Without their belief in me, I would not have been able to achieve all that I have.

Additionally, I would like to express my deep gratitude to my wife, Ting Hu, for her unwavering support and encouragement. Her faith in me has been the foundation upon which my academic achievements have been built. I am incredibly grateful to her for agreeing to my decision to resign from my well-paid job and pursue my Ph.D. Thank you for believing in me in pursuing my dreams. Furthermore, I am grateful for the sacrifices my wife has made for our partnership. After completing her master's degree, she gave

up the opportunity to pursue a career in a nice hospital and moved to the Netherlands to be with me. Her selflessness and love have been a constant source of strength and motivation for me, and I feel incredibly fortunate to have her by my side every day. I could not have made it this far without her. I am forever appreciative for her presence in my life.

Next, I would like to thank my Ph.D. committee members, Prof.dr. Johan Smit, Prof.dr. ir. J.A. La Poutré, Prof.dr.ir. P. Bauer, Dr. P.H. Nguyen, Dr. P. Mohajerin Esfahani, Dr. S. Chatzivasileiadis, and Dr.ir. S.E. Verwer for their assessment of this thesis.

I would like to express my gratitude to my coauthors Kaikai Pan, Ensieh Sharifnia, and Zhi Gao for their invaluable contributions to our academic work. Their insights and expertise have greatly enhanced the quality of our research, and I am deeply grateful for their collaboration. I would like to thank Dr. Milos Cvetkovic for his academic suggestions at the beginning of my Ph.D. journey. Many thanks to Umer, Digvijay and Hazem for their help. We had so many happy moments together. Moreover, I would like to express my appreciation to Ellen, Sharmila, and Carla for their invaluable assistance, as well as Remoko for his outstanding technical support.

I would like to express my gratitude to my Chinese friends Aihui, Le Liu, Yigu, Shengren, Haiwei, Na Li, Siyuan, and Lian Liu for their invaluable contributions to my academic and social experiences. Our stimulating discussions and shared activities have left me with many cherished memories, and I am grateful for their friendship and support. I also want to thank my current colleagues, Dr. Jochen Cremer, Dr. Pedro P. Vergara, Dr. Alex Stefanov, Dr.ir. Marjan Popov, Dr.ir. Jose Rueda Torres, Dr. Aleksandra Lekic, Qisong, Nanda, Wouter, Kutay, Ajay, Vetrivel, Alfan, Amirreza, Mojtaba, Roman, Ties, Zeynab, Demetris, Ali, Mert, Neda, Nidarshan, Dong Liu, Fan Xie, Shen Yan, Hongjin, Shuyi, Weijie, Nan Lin, Aleksandar, Ioannis, Raifa, Lucas, and Farzad, as well as my former colleagues Rishabh, Zhou Liu, Haiyan, Da Wang, Swasti, Arun, Arcadio, Claudio, Roland, Ilya, Nakis, and Matija. Your diverse research interests have broadened my horizons and deepened my understanding of various topics. Working alongside individuals from different cultures has shown me the beauty and richness of our differences, and I am grateful for the opportunity to learn from you all.

I am grateful for the invaluable friendship of my high school classmate Haorui Peng and my middle school classmate Linyu Lu. Their delicious home-cooked meals and lively conversations in our local dialects have brought me comfort and joy, reminding me of the warmth and familiarity of home. Our travels together have created unforget-

table memories filled with happiness. Thank you for making my time in the Netherlands all the more meaningful and unforgettable. Many thanks also to Xingchen, Gaoyang, Yuexiang, and Jiawang for their help, friendship, and care during my Ph.D. journey.